Uniform Bahadur Representation for Local Polynomial Estimates of M-Regression and Its Application to The Additive Model

Efang Kong*
EURANDOM, The Netherlands
Oliver Linton†
London School of Economics, UK
Yingcun Xia‡
National University of Singapore, Singapore

Summary

We use local polynomial fitting to estimate the nonparametric M-regression function for strongly mixing stationary processes $\{(Y_i, \underline{X}_i)\}$. We establish a strong uniform consistency rate for the Bahadur representation of estimators of the regression function and its derivatives. These results are fundamental for statistical inference and for applications that involve plugging in such estimators into other functionals where some control over higher order terms are required. We apply our results to the estimation of an additive M-regression model.

Key words: Additive model; Bahadur representation; Local polynomial fitting; M-regression; Strongly mixing processes; Uniform strong consistency.

1 Introduction

In many contexts one wants to evaluate the properties of some procedure that is a functional of some given estimators. It is useful to be able to work with some plausible high level assumptions about those estimators rather than to rederive their properties for each different application. In a fully parametric (and stationary, weakly dependent data) context it is quite common to

*Eurandom, Technische Universiteit Eindhoven, The Netherlands. E-mail address: `kong@eurandom.tue.nl`.
†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. http://econ.lse.ac.uk/staff/olinton/~index_own.html. E-mail address: `o.linton@lse.ac.uk`.
‡Department of Statistics and Applied Probability, National University of Singapore, Singapore. http://www.stat.nus.edu.sg/~staxyc. E-mail address: `staxyc@nus.edu.sg`.

assume that estimators are root-n consistent and asymptotically normal. In some cases this property suffices; in other cases one needs to be more explicit in terms of the linear expansion of these estimators, but in any case such expansions are quite natural and widely applicable. In a nonparametric context there is less agreement about the use of such expansions and one often sees standard properties of standard estimators derived anew for a different purpose. It is our objective to provide results that can circumvent this. The types of application we have in mind are estimation of semiparametric models where the parameters of interest are explicit or implicit functionals of nonparametric regression functions and their derivatives; see Powell (1994), Andrews (1994) and Chen, Linton and Van Keilegom (2003). Another class of applications includes estimation of structured nonparametric models like additive models (Linton and Nielsen, 1995) or generalized additive models (Linton, Sperlich and Van Keilegom, 2007).

We motivate our results in a simple i.i.d. setting. Suppose we have a random sample $\{Y_i, X_i\}_{i=1}^n$ and consider the Nadaraya-Watson estimator of the regression function $m(x) = E(Y_i|X_i = x)$,

$$\hat{m}(x) = \frac{\hat{r}(x)}{\hat{f}(x)} = \frac{n^{-1}\sum_{i=1}^n K_h(X_i - x)Y_i}{n^{-1}\sum_{i=1}^n K_h(X_i - x)},$$

where K is a symmetric density function, h is a bandwidth and $K_h(.) = K(./h)/h$. Standard arguments (Härdle, 1990) show that under suitable smoothness conditions

$$\hat{m}(x) - m(x) = h^2 b(x) + \frac{1}{nf(x)}\sum_{i=1}^n K_h(X_i - x)\varepsilon_i + R_n(x), \tag{1}$$

where $b(x) = \int u^2 K(u)du[m''(x) + 2m'(x)f'(x)/f(x)]/2$, while $f(x)$ is the covariate density and $\varepsilon_i \equiv Y_i - m(X_i)$ is the error term. The remainder term $R_n(x)$ is of smaller order (almost surely) than the two leading terms. Such an expansion is sufficient to derive the central limit theorem for $\hat{m}(x)$ itself, but generally is not sufficient if $\hat{m}(x)$ is to be plugged into some semiparametric procedure. For example, suppose we estimate the parameter $\theta_0 = \int m(x)^2 dx$ by $\hat{\theta} = \int \hat{m}(x)^2 dx$, where the integral is over some compact set $\mathcal{D}$; we would expect to find that $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal. Based on expansion (1), the argument goes like this. First, we obtain

the following

$$n^{1/2}(\hat{\theta}-\theta_0) = 2n^{1/2}\int m(x)\{\hat{m}(x)-m(x)\}dx + n^{1/2}\int [\hat{m}(x)-m(x)]^2 dx.$$

If it can be shown that $\hat{m}(x)-m(x) = o(n^{-1/4})$ a.s. uniformly in $x \in \mathcal{D}$ (such results are widely available; see for example Masry (1996)), we have

$$n^{1/2}(\hat{\theta}-\theta_0) = 2n^{1/2}\int m(x)\{\hat{m}(x)-m(x)\}dx + o(1), \quad a.s.$$

Note that the quantity on the right hand side is the term in assumption 2.6 of Chen, Linton, and Van Keilegom (2003) which is assumed to be asymptotically normal. It is the verification of this condition with which we are now concerned. If we substitute in the expansion (1) we obtain

$$\begin{aligned} n^{1/2}(\hat{\theta}-\theta_0) = 2n^{1/2}h^2\int m(x)b(x)dx + 2n^{1/2}\int \frac{m(x)}{f(x)} n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i dx \\ + 2n^{1/2}\int m(x)R_n(x)dx + o(1), \quad a.s. \end{aligned}$$

If $nh^4 \to 0$, then the first term (the smoothing bias term) is $o(1)$. By a change of variable, the second term (the stochastic term) can be written as a sum of independent random variables with zero mean

$$\begin{aligned} n^{1/2}\textstyle\int m(x)f^{-1}(x)n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i dx = n^{-1/2}\sum_{i=1}^{n}\xi_n(X_i)\varepsilon_i, \\ \xi_n(X_i) = \textstyle\int m(X_i + uh)f^{-1}(X_i + uh)K(u)du, \end{aligned}$$

and this term obeys the Lindeberg central limit theorem under standard conditions. The problem is that equation (1) only guarantees that $\int m(x)R_n(x)dx = o(n^{-2/5})$ a.s. at best. Actually, in this case it is possible to derive a more useful Bahadur expansion (Bahadur, 1966) for the kernel estimator

$$\hat{m}(x) - m(x) = h^2 b_n(x) + \{E\hat{f}(x)\}^{-1}n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i + R_n^*(x), \tag{2}$$

where $b_n(x)$ is deterministic and satisfies $b_n(x) \to b(x)$ uniformly in $x \in \mathcal{D}$, and $E\hat{f}(x) \to f(x)$ uniformly in $x \in \mathcal{D}$, while the remainder term now satisfies

$$\sup_{x\in\mathcal{D}} |R_n^*(x)| = O\left(\frac{\log n}{nh}\right) \quad a.s. \tag{3}$$

This property is a consequence of the uniform convergence rate of $\hat{f}(x) - E\hat{f}(x)$, $n^{-1}\sum_{i=1}^{n} K_h(x - X_i)\{m(X_i) - m(x)\} - EK_h(X_i - x)\{m(X_i) - m(x)\}$ and $n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i$ that follow from, for example Masry (1996). Clearly, by appropriate choice of h, $R_n^*(x)$ can be made to be $o(n^{-1/2})$ $a.s.$ uniformly over $\mathcal{D}$ and thus $2n^{1/2}\int m(x)R_n^*(x)dx = o(1)$ $a.s.$. Therefore, to derive asymptotic normality for $n^{1/2}(\hat{\theta} - \theta_0)$, one can just work with the two leading terms in (2). These terms are slightly more complicated than in the previous expansion but are still sufficiently simple for many purposes; in particular, $b_n(x)$ is uniformly bounded so that provided $nh^4 \to 0$, the smoothing bias term satisfies $h^2 n^{1/2}\int m(x)b_n(x)dx \to 0$, while the stochastic term is a sum of zero mean independent random variables

$$n^{1/2}\int \frac{m(x)}{\overline{f}(x)} n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i dx = n^{-1/2}\sum_{i=1}^{n} \overline{\xi}_n(X_i)\varepsilon_i$$

$$\overline{\xi}_n(X_i) = \int \frac{m(X_i + uh)}{\overline{f}(X_i + uh)} K(u)du,$$

and obeys the Lindeberg central limit theorem under standard conditions, where $\overline{f}(x) = E\hat{f}(x)$. This argument shows the utility of the Bahadur expansion (2). There are many other applications of this result because a host of probabilistic results are available for random variables like $n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\varepsilon_i$ and integrals thereof.

The one-dimensional Nadaraya-Watson estimator for i.i.d. data is particularly easy to analyze and the above arguments are well known. However, the limitations of this estimator are manyfold and there are good theoretical reasons for working instead with the local polynomial class of estimators (Fan and Gijbels, 1996). In addition, for many data especially financial time series data one may have concerns about heavy tails or outliers that point in the direction of using robust estimators like the local median or local quantile method, perhaps combined with local polynomial fitting. We examine a general class of (nonlinear) M-regression function (that is, location functionals defined through minimization of a general objective function $\rho(.)$) and derivative estimators. We treat a general time series setting where the multivariate data are strongly mixing. Under mild conditions, we establish a uniform strong Bahadur expansion like (2) and (3) with remainder term of order $(\log n/nh^d)^{3/4}$ almost surely, which is almost optimal

or in other words can't be improved further based on the results in Kiefer (1967) under i.i.d. setting. The leading terms are linear and functionals of them can be analyzed simply. The remainder term can be made to be $o(n^{-1/2})$ a.s. under restrictions on the dimensionality in relation to the amount of smoothness possessed by the M-regression function.

The best convergence rate of unrestricted nonparametric estimators strongly depends on d, the dimension of $\underline{x}$. The rate decreases dramatically as d increases (Stone, 1982). This phenomenon is the so-called "curse of dimensionality". One approach to reduce the curse is by imposing model structure. A popular model structure is the additive model assuming that

$$m(x_1, \ldots, x_d) = c + m_1(x_1) + \ldots + m_d(x_d), \tag{4}$$

where c is an unknown constant and $m_k(.),\ k = 1, \ldots, d$ are unknown functions which have been normalized such that $Em_k(\mathbf{x}_k) = 0$ for $k = 1, \ldots, d$. In this case, the optimal rate of convergence is the same as in univariate nonparametric regression (Stone, 1986). An additive M-regression function is given by (4) with $m(x)$ being the M-regression function defined in (5). Previous work on additive quantile regression, for example, includes Linton (2001) and Horowitz and Lee (2005) for the i.i.d. case. An interesting application of the additive M-regression model is to combine (4) with the volatility model

$$Y_i = \sigma_i \varepsilon_i \quad \text{and} \quad \ln \sigma_i^2 = m(X_i),$$

where $X_i = (Y_{i-1}, \ldots, Y_{i-d})^\top$. We suppose that ε_i satisfies $E[\varphi(\ln \varepsilon_i^2; 0)|X_i] = 0$ for some function $\varphi(.)$, whence $m(.)$ is the conditional M-regression of $\ln Y_i^2$ given X_i. Peng and Yao (2003) have applied LAD estimation to parametric ARCH and GARCH models and have shown the superior robustness property of this procedure over Gaussian QMLE with regard to heavy tailed innovations. This heavy tail issue also arises in nonparametric regression models, which is why our procedures may be useful. Empirical evidence also suggest that moderately high frequency financial data are often heavy tailed. We apply our Bahadur expansions to the study of marginal integration estimators (Linton and Nielsen, 1995) of the component functions in additive M-regression model in which case we only need the remainder term to be $o(n^{-p/(2p+1)})$ a.s., where

p is a smoothness index.

Bahadur representations (Bahadur, 1966) have been widely studied and applied, with notable refinements in the i.i.d. setting by Kiefer (1967). A recent paper of Wu (2005) extends these results to a general class of dependent processes and provides a review. The closest paper to ours is Hong (2003) who established a Bahadur representation for essentially the same local polynomial M-regression estimator as ours. However, his results are: (a) pointwise, i.e., for a single x only; (b) the covariates are univariate; (c) for i.i.d. data. Clearly, this limits the range of applicability of his results, and specifically, the applications to semiparametric or additive models are perforce precluded.

2 The General Setting

Let $\{(Y_i, \underline{X}_i)\}$ be a jointly stationary processes, where $\underline{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{id})^\top$ with $d \geq 1$ and Y_i is a scalar. As dependent observations are considered in this paper, we introduce here the mixing coefficient. Let $\mathbf{F}_s^t$ be the $\sigma-$ algebra of events generated by random variables $\{(Y_i, \underline{X}_i), s \leq i \leq t\}$. A stationary stochastic processes $\{(Y_i, \underline{X}_i)\}$ is strongly mixing if

$$\sup_{\substack{A \in \mathbf{F}_{-\infty}^0 \\ B \in \mathbf{F}_k^\infty}} |P[AB] - P[A]P[B]| = \gamma[k] \to 0, \text{ as } k \to \infty,$$

and $\gamma[k]$ is called the strong mixing coefficient.

Suppose $\rho(.;.)$ is a loss function. Our first goal is to estimate the multivariate M-regression function

$$m(x_1, \cdots, x_d) = \arg\min_\theta E\{\rho(Y_i; \theta) | \underline{X}_i = (x_1, \cdots, x_d)\}, \tag{5}$$

and its partial derivatives based on observations $\{(Y_i, \underline{X}_i)\}_{i=1}^n$. An important example of the M-function is the $q-$th $(0 < q < 1)$ quantile of Y_i given $\underline{X}_i = (x_1, \cdots, x_d)^\top$, with loss function given by $\rho(y; \theta) = (2q-1)(y-\theta) + |y-\theta|$. Another example is the L_q criterion $\rho(y; \theta) = |y-\theta|^q$ for $q > 1$, which includes the least square criterion $\rho(y; \theta) = (y-\theta)^2$ with $m(.)$ the conditional expectation of Y_i given $\underline{X}_i$. Yet another example is the celebrated Huber's function (Huber,

1973)

$$\rho(t) = t^2/2I\{|t| < k\} + (k|t| - k^2/2)I\{|t| \geq k\}. \tag{6}$$

Suppose $m(\underline{x})$ is differentiable up to order $p+1$ at $\underline{x} = (x_1, ..., x_d)^\top$. Then the multivariate p'th order local polynomial approximation of $m(\underline{z})$ for any $\underline{z}$ close to $\underline{x}$ is given by

$$m(\underline{z}) = \sum_{0 \leq |\underline{r}| \leq p} \frac{1}{\underline{r}!} D^{\underline{r}} m(\underline{x})(\underline{z} - \underline{x})^{\underline{r}},$$

where $\underline{r} = (r_1, ..., r_d)$, $|\underline{r}| = \sum_{i=1}^d r_i$, $\underline{r}! = r_1! \times \cdots \times r_d!$ and

$$D^{\underline{r}} m(\underline{x}) = \frac{\partial^{\underline{r}} m(\underline{x})}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}, \quad \underline{x}^{\underline{r}} = x_1^{r_1} \times ... \times x_d^{r_d}, \quad \sum_{0 \leq |\underline{r}| \leq p} = \sum_{j=0}^{p} \sum_{r_1=0}^{j} ... \sum_{\substack{r_d=0 \\ r_1+...+r_d=j}}^{j}. \tag{7}$$

Let $K(\underline{u})$ be a density function on R^d, h a bandwidth and $K_h(u) = K(u/h)$. With observations $\{(Y_i, \underline{X}_i)\}_{i=1}^n$, we consider minimizing the following quantity with respect to $\beta_{\underline{r}}$, $0 \leq |\underline{r}| \leq p$

$$\sum_{i=1}^n K_h(\underline{X}_i - \underline{x})\rho\Big(Y_i; \sum_{0 \leq |\underline{r}| \leq p} \beta_{\underline{r}}(\underline{X}_i - \underline{x})^{\underline{r}}\Big). \tag{8}$$

Denote by $\hat{\beta}_{\underline{r}}(\underline{x})$, $0 \leq |r| \leq p$, the minima of (8). The M-function $m(\underline{x})$ and its derivatives $D^{\underline{r}} m(\underline{x})$ are then estimated respectively by

$$\hat{m}(\underline{x}) = \hat{\beta}_{\underline{0}}(\underline{x}) \quad \text{and} \quad \hat{D}^{\underline{r}} m(\underline{x}) = \underline{r}! \hat{\beta}_{\underline{r}}(\underline{x}), \ 1 \leq |\underline{r}| \leq p. \tag{9}$$

3 Main Results

In Theorem 3.2 below we give our main result, the uniform strong Bahadur representation for the vector $\hat{\beta}_p(\underline{x})$. We first need to develop some notations to define the leading terms in the expansion.

Let $N_i = \binom{i+d-1}{d-1}$ be the number of distinct $d-$tuples $\underline{r}$ with $|\underline{r}| = i$. Arrange these $d-$tuples as a sequence in a lexicographical order(with the highest priority given to the last position so that $(0, \cdots, 0, i)$ is the first element in the sequence and $(i, 0, \cdots, 0)$ the last element). Let τ_i denote this 1-to-1 mapping, i.e. $\tau_i(1) = (0, \cdots, 0, i), \cdots, \tau_i(N_i) = (i, 0, \cdots, 0)$. For each

$i = 1, \cdots, p$, define a $N_i \times 1$ vector $\mu_i(\underline{x})$ with its kth element given by $\underline{x}^{\tau_i(k)}$ and write $\mu(\underline{x}) = (1, \mu_1(\underline{x})^\top, \cdots, \mu_p(\underline{x})^\top)^\top$, which is a column vector of length $N = \sum_{i=0}^{p} N_i$. Similarly define vectors $\beta_p(\underline{x})$ and $\underline{\beta}$ through the same lexicographical arrangement of $D^{\underline{r}}m(\underline{x})$ and $\beta_{\underline{r}}$ in (8) for $0 \leq |\underline{r}| \leq p$. Thus (8) can be rewritten as

$$\sum_{i=1}^{n} K_h(\underline{X}_i - \underline{x})\rho(Y_i; \mu(\underline{X}_i - \underline{x})^\top \underline{\beta}). \tag{10}$$

Suppose the minimizer of (10) is denoted as $\tilde{\beta}_n(\underline{x})$. Let $\hat{\beta}_p(\underline{x}) = W_p\hat{\beta}_n(\underline{x})$, where W_p is the diagonal matrix with diagonal entries the lexicographical arrangement of $\underline{r}!,\ 0 \leq |\underline{r}| \leq p$.

Let $\nu_{\underline{i}} = \int K(\underline{u})\underline{u}^{\underline{i}}d\underline{u}$. For $g(.)$ given in (24), define

$$\nu_{n\underline{i}}(\underline{x}) = \int K(\underline{u})\underline{u}^{\underline{i}}g(\underline{x} + h\underline{u})f(\underline{x} + h\underline{u})d\underline{u}.$$

For $0 \leq j, k \leq p$, let $S_{j,k}$ and $S_{n,j,k}(\underline{x})$ be two $N_j \times N_k$ matrices with their (l, m) elements respectively given by

$$\Big[S_{j,k}\Big]_{l,m} = \nu_{\tau_j(l)+\tau_k(m)}(\underline{x}), \quad \Big[S_{n,j,k}(\underline{x})\Big]_{l,m} = \nu_{n,\tau_j(l)+\tau_k(m)}(\underline{x}). \tag{11}$$

Now define the $N \times N$ matrices S_p and $S_{n,p}(\underline{x})$ by

$$S_p = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\ \vdots & \ddots & \vdots & \\ S_{p,0} & S_{p,1} & \cdots & S_{p,p} \end{bmatrix}, \quad S_{n,p}(\underline{x}) = \begin{bmatrix} S_{n,0,0}(\underline{x}) & S_{n,0,1}(\underline{x}) & \cdots & S_{n,0,p}(\underline{x}) \\ S_{n,1,0}(\underline{x}) & S_{n,1,1}(\underline{x}) & \cdots & S_{n,1,p}(\underline{x}) \\ \vdots & \ddots & \vdots & \\ S_{n,p,0}(\underline{x}) & S_{n,p,1}(\underline{x}) & \cdots & S_{n,p,p}(\underline{x}) \end{bmatrix}.$$

According to Lemma 5.8, $S_{n,p}(\underline{x})$ converges to $g(\underline{x})f(\underline{x})S_p$ uniformly in $\underline{x} \in \mathcal{D}$ almost surely. Hence for $|S_p| \neq 0$, we can define

$$\beta_n^*(\underline{x}) = -\frac{1}{nh^d}W_pS_{n,p}^{-1}(\underline{x})H_n^{-1}\sum_{i=1}^{n} K_h(\underline{X}_i - \underline{x})\varphi(Y_i, \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x}))\mu(\underline{X}_i - \underline{x}), \tag{12}$$

where $\varphi(.;.)$ is the piecewise derivative of $\rho(.,.)$, as defined in (A1) and H_n is the diagonal matrix with diagonal entries $h^{|\underline{r}|},\ 0 \leq |\underline{r}| \leq p$ in the aforementioned lexicographical order. The quantity $\beta_n^*(\underline{x})$ is the leading term of our expansion; it contains both a bias term, $E\beta_n^*(\underline{x})$, and a stochastic leading term $\beta_n^*(\underline{x}) - E\beta_n^*(\underline{x})$.

Denote the typical element of $\beta_n^*(\underline{x})$ by $\beta_{n\underline{r}}^*(\underline{x}),\ 0 \le |\underline{r}| \le p$ and the density function of $\underline{X}$ by $f(.)$. The following results on $E\beta_{n\underline{r}}^*(\underline{x})$ is an extension of Proposition 2.2 in Hong (2003) to the multivariate case.

Proposition 3.1 *If $f(\underline{x}) > 0$ and conditions (A1)-(A5) in the Appendix hold, then*

$$E\beta_{n\underline{r}}^*(\underline{x}) = \begin{cases} -h^{p+1} e_{N(\underline{r})} W_p S_p^{-1} B_1 \mathbf{m}_{p+1}(\underline{x}) + o(h^{p+1}), & \text{for } p - |\underline{r}| \text{ odd,} \\ \\ -h^{p+2} e_{N(\underline{r})} W_p S_p^{-1} \Big[\{fg\}^{-1}(\underline{x}) \mathbf{m}_{p+1}(\underline{x}) \{\tilde{M}(\underline{x}) - N_p S_p^{-1} B_1\} + B_2 \mathbf{m}_{p+2}(\underline{x}) \Big] & \\ +o(h^{p+2}), & \text{for } p - |\underline{r}| \text{ even,} \end{cases}$$

where $N(\underline{r}) = \tau_{|\underline{r}|}^{-1}(\underline{r}) + \sum_{k=0}^{|\underline{r}|-1} N_k$, e_i is a $N \times 1$ vector having 1 as the ith entry with all other entries 0, and $B_1 = [S_{0,p+1}, S_{1,p+1}, \cdots S_{p,p+1}]^\top$, $B_2 = [S_{0,p+2}, S_{1,p+2}, \cdots S_{p,p+2}]^\top$.

We next present our main result, the Bahadur representation for local polynomial estimates $\hat{\beta}_p(\underline{x})$.

Theorem 3.2 *Suppose (A1)-(A7) in the Appendix hold with $\lambda_2 = (p+1)/2(p+s+1)$ for some $s \ge 0$ and $\mathcal{D}$ is any compact subset of R^d. Then*

$$\sup_{\underline{x} \in \mathcal{D}} |H_n\{\hat{\beta}_p(\underline{x}) - \beta_p(\underline{x})\} - \beta_n^*(\underline{x})| = O\Big(\Big\{\frac{\log n}{nh^d}\Big\}^{\lambda(s)}\Big) \text{ almost surely,}$$

where $|.|$ is taken to be the sup norm and

$$\lambda(s) = \min\Big\{\frac{p+1}{p+s+1}, \frac{3p+3+2s}{4p+4s+4}\Big\}.$$

Remark 1. According to Theorem 1 in Kiefer (1967), the point-wise sharpest bound of the remainder term in Bahadur representation of the sample quantiles is $(\log\log n/n)^{3/4}$. As $\lambda(0) = 3/4$, we could safely claim the results here could not be further improved for a general class of loss functions $\rho(.)$ specified by (A1) and (A2). Nevertheless, it is possible to derive stronger results, if the concerned loss functions enjoy higher degree of smoothness; see (3) in which case $\rho(.)$ is the squared loss function. More specifically, suppose $\varphi(.)$ is Lipschitz continuous and (A1)-(A7) in the Appendix hold with $\lambda_2 = 1/2$ and $\lambda_1 = 1$. Then we prove in the Appendix that with probability 1 and uniformly in $\underline{x} \in \mathcal{D}$,

$$\sup_{\underline{x} \in \mathcal{D}} |H_n\{\hat{\beta}_p(\underline{x}) - \beta_p(\underline{x})\} - \beta_n^*(\underline{x})| = O\Big(\frac{\log n}{nh^d}\Big) \text{ almost surely.} \tag{13}$$

Remark 2. The dependence among the observations doesn't have any impact on the rate of uniform convergence, given that the degree of the dependence, as measured by the mixing coefficient $\gamma[k]$, is weak enough such that (20) and (21) are satisfied. This is in accordance with the results in Masry (1996), where he proved that for local polynomial estimator of the conditional mean function, the uniform convergence rate is $(nh^d/\log n)^{-1/2}$, the same as in the independent case.

Remark 3. It is of practical interest to provide an explicit rate of decay for the strong mixing coefficient $\gamma[k]$ of the form $\gamma[k] = O(1/k^c)$ for some $c > 0$ (to be determined) for Theorem 3.2 to hold. It is easily seen that, among all the conditions imposed on $\gamma[k]$, the summability condition (21) is the most restrictive. We assume that

$$h = h_n \sim (\log n/n)^{\bar{a}} \text{ for some } \frac{1}{2(p+s+1)+d} \leq \bar{a} < \frac{1}{d}\Big\{1 - \frac{4}{(1-\lambda_2)\nu_2 - 4\lambda_1 + 2(1+\lambda_2)}\Big\}$$

whence (19) is satisfied. Algebraic calculations show that the summability condition (21) is satisfied provided that

$$c > \nu_2 \frac{(1-\bar{a}d)\{(1-\lambda_2)(4N+1) + 8N\lambda_1\} + 10 + (4+8N)\bar{a}d}{2(1-\lambda_2)(1-\bar{a}d)\nu_2 - 8\bar{a}d + 4(1-\bar{a}d)(1-\lambda_2-2\lambda_1)} - 1 \equiv c(d,p,\nu_2,\bar{a},\lambda_1,\lambda_2). \quad (14)$$

Note that we would need the following condition

$$\nu_2 > 2 + \frac{4\{\bar{a}d + (1-\bar{a}d)\lambda_1\}}{(1-\bar{a}d)(1-\lambda_2)}$$

to secure positive denominator for (14). As $c(d,p,\nu_2,\bar{a},\,\lambda_1,\lambda_2)$ is decreasing in $\nu_2 (\leq \nu_1)$, there is a tradeoff between the order ν_1 of the moment $E|\varphi(\varepsilon_i)|^{\nu_1} < \infty$ and the decay rate of the strong mixing coefficient $\gamma[k]$: the existence of higher order moments allows $\gamma[k]$ to decay more slowly.

Remark 4. It is trivial to generalize the result in Theorem 3.2 to functionals of the M-estimates $\hat{\beta}_p(\underline{x})$. Denote the typical elements of $\hat{\beta}_p(\underline{x})$ and $\beta_p(\underline{x})$ by $\hat{\beta}_{p\underline{r}}(\underline{x})$ and $\beta_{p\underline{r}}(\underline{x})$, $0 \leq |\underline{r}| \leq p$ respectively. Suppose $G(.):\ R^d \to R$ satisfies that for any compact set $\mathcal{D} \subset R^d$, there exists some constant $C > 0$, such that $|G',(\beta_{p\underline{r}}(\underline{x}))| \leq C$ and $|G''(\beta_{p\underline{r}}(\underline{x}))| \leq C$ for all $\underline{x} \in \mathcal{D}$. Then with probability 1,

$$\sup_{\underline{x}\in\mathcal{D}} \Big| h^{|\underline{r}|} \Big[G\{\hat{\beta}_{p\underline{r}}(\underline{x})\} - G\{\beta_{p\underline{r}}(\underline{x})\Big] - G'\{\beta_{p\underline{r}}(\underline{x})\}\beta^*_{n\underline{r}}(\underline{x}) \Big| = O\Big(\Big\{\frac{\log n}{nh^d}\Big\}^{\lambda(s)}\Big) \quad (15)$$

uniformly for all $\underline{x} \in \mathcal{D}$.

The following proposition follows from Theorem 3.2 and uniform convergence of sum of weakly dependent zero mean random variables.

Corollary 3.3 *Suppose conditions in Theorem 3.2 hold with $s = 0$. Then with probability 1 we have, uniformly in $\underline{x} \in \mathcal{D}$,*

$$H_n\{\hat{\beta}_p(\underline{x}) - \beta_p(\underline{x})\} - E\beta_n^*(\underline{x}) - \frac{W_p H_n^{-1}}{nh^d} S_{np}^{-1}(\underline{x}) \sum_{i=1}^n K_h(\underline{X}_i - \underline{x})\varphi(\varepsilon_i)\mu(\underline{X}_i - \underline{x}) = O\Big(\Big\{\frac{\log n}{nh^d}\Big\}^{3/4}\Big).$$

4 M-Estimation of the Additive model

In this section, we apply our main result to derive the properties of a class of estimators in the additive M-regression model (4). In terms of estimating the component functions $m_k(.)$, $k = 1, \ldots, d$ in (4), the marginal integration method (Linton and Nielsen, 1995) is known to achieve the optimal rate under certain conditions. This involves estimating first the unrestricted M-regression function $m(.)$ and then integrating it over some directions. Partition $\underline{X}_i = (x_1, \ldots, x_d)$ as $\underline{X}_i = (\mathbf{x}_{1i}, \underline{X}_{2i})$, where $\mathbf{x}_{1i}$ is the one dimensional direction of interest and $\underline{X}_{2i}$ is a $d-1$ dimensional nuisance direction. Let $\underline{x} = (x_1, \underline{x}_2)$ and define the functional

$$\phi_1(x_1) = \int m(x_1, \underline{x}_2) f_2(\underline{x}_2) d\underline{x}_2, \tag{16}$$

where $f_2(\underline{x}_2)$ is the joint density of $\underline{X}_{2i}$. Under the additive structure (4), $\phi_1(.)$ is $m_1(.)$ up to a constant. Replace $m(.)$ in (16) with $\hat{\beta}_0(x_1, \underline{x}_2) \equiv \hat{\beta}_{\underline{0}}(\underline{x})$ given by (9) and $\phi_1(x_1)$ can thus be estimated by the sample version of (16):

$$\phi_{n1}(x_1) = n^{-1} \sum_{i=1}^n \hat{\beta}_0(x_1, \underline{X}_{2i}).$$

As noted by Linton and Härdle (1996) and Hengartner and Sperlich (2005), cautious choice of the bandwidth is crucial for $\phi_{n1}(.)$ to be asymptotically normal. They suggested different bandwidths be engaged for the direction of interest X_1 and the $d-1$ dimensional nuisance direction $\underline{X}_2$, say h_1 and h respectively. Sperlich et al (1998) and Linton et al (1999) provide

an extensive study of the small sample properties of marginal integration estimators, including an evaluation of bandwidth choice.

The following corollary is about the asymptotic properties of $\phi_{n1}(.)$.

Corollary 4.1 *Suppose the support of $\underline{X}$ is $[0,1]^{\otimes d}$ with strictly positive density function. Let the conditions in Proposition 3.3 hold with $\mathrm{T}_n \equiv \{r(n)/\min(h_1,h)\}^d$ and the h^d replaced by $h_1 h^{d-1}$ in all the notations defined either in (18) or (19). If $h_1 \propto n^{-1/(2p+3)}$, $h = O(h_1)$ and (19) is modified as*

$$nh_1 h^{3(d-1)}/\log^3 n \to \infty,\ n^{-1}\{r(n)\}^{\nu_2/2} d_n \log n / M_n^{(2)} \to \infty. \tag{17}$$

Then we have

$$(nh_1)^{1/2}\{\phi_{n1}(x_1) - \phi_1(x_1)\} \xrightarrow{L} N(e_1 W_p S_p^{-1} B_1 E\mathbf{m}_{p+1}(x_1, \underline{X}_2), \tilde{\sigma}^2(x_1)),$$

where '$\xrightarrow{L}$' stands for convergence in distribution,

$$\tilde{\sigma}^2(x_1) = \Big\{ \int_{[0,1]^{\otimes d-1}} \{fg^2\}^{-1}(x_1, \underline{X}_2) f_2^2(\underline{X}_2) \sigma^2(x_1, \underline{X}_2) d\underline{X}_2 \Big\} e_1 S_p^{-1} K_2 K_2^\top S_p^{-1} e_1^\top,$$

$\sigma^2(\underline{x}) = E[\varphi^2(\varepsilon)|\underline{X} = \underline{x}]$ and $K_2 = \int_{[0,1]^{\otimes d}} K(\underline{v})\mu(\underline{v})d\underline{v}$. In particular for additive quantile regression, i.e. $\rho(y;\theta) = (2q-1)(y-\theta) + |y-\theta|$, we have

$$\tilde{\sigma}^2(x_1) = q(1-q)\Big\{ \int_{[0,1]^{\otimes d-1}} f^{-1}(x_1, \underline{X}_2) f_\varepsilon^{-2}(0|x_1, \underline{X}_2) f_2^2(\underline{X}_2) d\underline{X}_2 \Big\} e_1 S_p^{-1} K_2 K_2^\top S_p^{-1} e_1^\top.$$

Remark 5. For conditions in Corollary 4.1 to hold, we would need $3d < 2p+5$, i.e. the order of local polynomial approximation increases as the dimension of the covariates $\underline{X}$ increases. See also the discussion in Hengartner and Sperlich (2005).

Remark 6. Besides asymptotic normality, we could also by applying Theorem 3.2 develop Bahadur representations for $\phi_{n1}(x_1)$, like those assumed in Linton, Sperlich and Van Keilegom (2007). Based on (15), similar results are also applicable to the generalized additive M-regression model where $G(m(x_1,\ldots,x_d)) = c + m_1(x_1) + \ldots + m_d(x_d)$ for some known smooth function $G(.)$, in which case the marginal integration estimator is given by the sample average of $G(\hat{m}(x_1, \underline{X}_{2i}))$.

5 Concluding Remarks

Our results can be useful in a variety of contexts including estimation of quite general nonlinear functionals of M-regression functions, and we have shown in one specific application how they can be applied.

Appendix: Regularity Conditions and Proofs

For any $M > 2$, $\lambda_2 \in (0, 1)$ and $\lambda_1 \in (\lambda_2, (1 + \lambda_2)/2]$, define

$$d_n = (nh^d/\log n)^{-(\lambda_1+\lambda_2/2)}(nh^d \log n)^{1/2},\ r(n) = (nh^d/\log n)^{(1-\lambda_2)/2}, \tag{18}$$

$$M_n^{(1)} = M(nh^d/\log n)^{-\lambda_1},\ M_n^{(2)} = M^{1/4}(nh^d/\log n)^{-\lambda_2},\ \mathrm{T}_n = \{r(n)/h\}^d$$

and L_n as the smallest integer such that $\log n(M/2)^{\mathrm{L}_n+1} > nM_n^{(2)}/d_n$. Let $\|.\|$ denote the Euclidean norm and C be a generic constant, which may have different values at each appearance. Let $\varepsilon_i \equiv Y_i - m(\underline{X}_i)$ and assume that the following conditions hold.

(A1) For each $y \in \mathcal{R}$, $\rho(y;\theta)$ is absolutely continuous in θ, *i.e.*, there is a function $\varphi(y;\theta) \equiv \varphi(y-\theta)$ such that for any $\theta \in \mathcal{R}$, $\rho(y;\theta) = \rho(y;0) + \int_0^\theta \varphi(y;t)dt$. The probability density function of ε_i is bounded, $E\{\varphi(\varepsilon_i)|\underline{X}_i\} = 0$ almost surely and $E|\varphi(\varepsilon_i)|^{\nu_1} < \infty$ for some $\nu_1 > 2$.

(A2) $\varphi(.)$ satisfies the Lipschitz condition in (a_j, a_{j+1}), $j = 0, \cdots, m$, where $a_1 < \cdots < a_m$ are the finite number of jump discontinuity points of $\varphi(.)$, $a_0 \equiv -\infty$ and $a_{m+1} \equiv +\infty$.

(A3) $K(.)$ has a compact support, say $[-1,1]^{\otimes d}$ and $|H_{\underline{j}}(\underline{u}) - H_{\underline{j}}(\underline{v})| \le C\|u - v\|$ for all j with $0 \le |\underline{j}| \le 2p+1$, where $H_{\underline{j}}(u) = \underline{u}^{\underline{j}}K(\underline{u})$.

(A4) The probability density function of $\underline{X}$, $f(.)$ is bounded and with bounded first order derivatives. The joint probability density of $(\underline{X}_0, \underline{X}_l)$ satisfies $f(\underline{u}, \underline{v}; l) \le C < \infty$ for all $l \ge 1$.

(A5) For $\underline{r}$ with $|\underline{r}| = p+1$, $D^{\underline{r}}m(\underline{x})$ is bounded with bounded first order derivative.

(A6) The bandwidth $h \to 0$ with

$$nh^d/\log n \to \infty,\ nh^{d+(p+1)/\lambda_2}/\log n < \infty,\ \ n^{-1}\{r(n)\}^{\nu_2/2} d_n \log n / M_n^{(2)} \to \infty, \qquad (19)$$

for some $2 < \nu_2 \leq \nu_1$ and the processes $\{(Y_i, \underline{X}_i)\}$ are strongly mixing with mixing coefficient $\gamma[k]$ satisfying

$$\sum_{k=1}^{\infty} k^a \{\gamma[k]\}^{1-2/\nu_2} < \infty \text{ for some } a > (p+d+1)(1-2/\nu_2)/d. \qquad (20)$$

Moreover, the bandwidth h and $\gamma[k]$ should jointly satisfy the following condition

$$\sum_{n=1}^{\infty} n^{3/2} \mathrm{T}_n \Big\{\frac{M_n^{(1)}}{d_n}\Big\}^{1/2} \frac{\gamma[r(n)(2^{\nu_2/2}/M)^{2\mathrm{L}_n/\nu_2}]}{r(n)(2^{\nu_2/2}/M)^{2\mathrm{L}_n/\nu_2}} \{4M^{2N}\}^{\mathrm{L}_n} < \infty,\ \forall M > 0. \qquad (21)$$

(A7) The conditional density $f_{\underline{X}|Y}$ of $\underline{X}$ given Y exists and is bounded. The conditional density $f_{(\underline{X}_1, \underline{X}_{l+1})|(Y_1, Y_{l+1})}$ of $(\underline{X}_1, \underline{X}_{l+1})$ given (Y_1, Y_{l+1}) exists and is bounded, for all $l \geq 1$.

Remark 7. Assumptions on $\varphi(.)$ in (A1) and (A2) are satisfied in almost all known robust and likelihood type regressions. For example, in $qth-$quantile regression, we have $\varphi(t) = 2qI\{t \geq 0\} + (2q-2)I\{t < 0\}$, while for the Huber's function (6), its piecewise derivative is given by

$$\varphi(t) = tI\{|t| < k\} + \mathrm{sign}(t)kI\{|t| \geq k\}.$$

Note that the condition $E\{\varphi(\varepsilon_i)|\underline{X}_i\} = 0$ *a.e.* is needed for model specification. Moreover, if the conditional density $f(y|\underline{x})$ of Y given $\underline{X}$ is also continuously differentiable with respect to y, then as proved in Hong (2003) there is a constant $C > 0$, such that for all small t and $\underline{x}$,

$$E\Big[\Big\{\varphi(Y; t+a) - \varphi(Y; a)\Big\}^2 |\underline{X} = \underline{u}\Big] \leq C|t| \qquad (22)$$

holds for all $(a, \underline{u})$ in a neighborhood of $(m(\underline{x}), \underline{x})$. Define

$$G(t, \underline{u}) = E\{\varphi(Y; t)|\underline{X} = \underline{u}\}, \quad G_i(t, \underline{u}) = (\partial^i/\partial t^i) G(t, \underline{u}),\ i = 1, 2, \qquad (23)$$

then it holds that

$$g(\underline{x}) = G_1(m(\underline{x}), \underline{x}) \geq C > 0,\ G_2(t, \underline{x}) \text{ bounded for all } \underline{x} \in \mathcal{D} \text{ and } t \text{ near } m(\underline{x}). \qquad (24)$$

Assumptions (A3)-(A7) are standard for nonparametric smoothing in multivariate time series analysis, see Masry (1996). For example, condition (20) is needed to bound the covariance of partial sums of time series as in Lemma 5.5, while (21) plays a similar role as (4.7b) in Masry (1996). It guarantees that the dependence of the time series is weakly enough such that the difference caused by the approximation of dependent random variables by independent ones (through Bradley's strong approximation theorem) is negligible; see Lemma 5.4. Of course, (21) is more stringent than (4.7b) in Masry (1996), which is due to the fact that the loss function $\rho(.)$ considered here is more general than the straightforward square loss.

Proof of Proposition 3.1. Write $\beta_n^*(\underline{x}) = -W_p S_{n,p}^{-1}(\underline{x}) \sum_{i=1}^n Z_{ni}(\underline{x})/n$, where

$$Z_{ni}(\underline{x}) = H_n^{-1} h^{-d} K_h(\underline{X}_i - \underline{x}) \varphi(Y_i, \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})) \mu(\underline{X}_i - \underline{x}).$$

We first focus on $EZ_{ni}(\underline{x})$. Based on (23) and (24), we have

$$\begin{aligned} E\{\varphi(Y_i, \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})) | \underline{X}_i\} &= G(\mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x}), \underline{X}_i) \\ &= -g(\underline{X}_i)\{m(\underline{X}_i) - \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})\} \\ &\quad + G_2(\xi_i(x), \underline{X}_i)\{m(\underline{X}_i) - \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})\}^2/2 \end{aligned}$$

for some $\xi_i(x)$ between $\mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})$ and $m(\underline{X}_i)$. Apparently, if $\underline{X}_i = \underline{x} + h\underline{v}$, then

$$m(\underline{X}_i) - \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x}) = h^{p+1} \sum_{|\underline{k}|=p+1} \frac{D^{\underline{r}} m(\underline{x})}{\underline{k}!} \underline{v}^{\underline{k}} + h^{p+2} \sum_{|\underline{k}|=p+2} \frac{D^{\underline{r}} m(\underline{x})}{\underline{k}!} \underline{v}^{\underline{k}} + o(h^{p+2}).$$

Therefore,

$$\begin{aligned} EZ_{ni}(\underline{x}) &= h^{p+1} \int K(\underline{v}) fg(\underline{x} + h\underline{v}) \mu(\underline{v}) \sum_{|\underline{k}|=p+1} \frac{D^{\underline{r}} m(\underline{x})}{\underline{k}!} \underline{v}^{\underline{k}} d\underline{v} \\ &\quad + h^{p+2} \int K(\underline{v}) fg(\underline{x} + h\underline{v}) \mu(\underline{v}) \sum_{|\underline{k}|=p+2} \frac{D^{\underline{r}} m(\underline{x})}{\underline{k}!} \underline{v}^{\underline{k}} d\underline{v} + o(h^{p+2}) \\ &\equiv T_1 + T_2. \end{aligned}$$

Now arrange the N_{p+1} elements of the derivatives $D^{\underline{r}} m(\underline{x})/\underline{r}!$ for $|\underline{r}| = p+1$ as a column vector $\mathbf{m}_{p+1}(\underline{x})$ using the lexicographical order introduced earlier and define $\mathbf{m}_{p+2}(\underline{x})$ in the similar

way. Let the $N \times N_{p+1}$ matrix B_{n1} and the $N \times N_{p+2}$ matrix B_{n2} be defined as

$$B_{n1}(\underline{x}) = \begin{bmatrix} S_{n,0,p+1}(\underline{x}) \\ S_{n,1,p+1}(\underline{x}) \\ \vdots \\ S_{n,p,p+1}(\underline{x}) \end{bmatrix}, \quad B_{n2}(\underline{x}) = \begin{bmatrix} S_{n,0,p+2}(\underline{x}) \\ S_{n,1,p+2}(\underline{x}) \\ \vdots \\ S_{n,p,p+2}(\underline{x}) \end{bmatrix},$$

where $S_{n,i,p+1}(\underline{x})$ and $S_{n,i,p+2}(\underline{x})$ is as given by (11). Therefore, $T_1 = h^{p+1} B_{n1}(\underline{x})\mathbf{m}_{p+1}(\underline{x})$, $T_2 = h^{p+2}B_{n2}(\underline{x})\mathbf{m}_{p+2}(\underline{x})$, and

$$E\beta_n^*(\underline{x}) = -W_p h^{p+1} S_{n,p}^{-1}(\underline{x}) B_{n1}(\underline{x})\mathbf{m}_{p+1}(\underline{x}) - W_p h^{p+2} S_{n,p}^{-1}(\underline{x}) B_{n2}(\underline{x})\mathbf{m}_{p+2}(\underline{x}) + o(h^{p+2}).$$

Let $\underline{e}_i, \ i = 1, \cdots, d$ be the $d \times 1$ vector having 1 in the ith entry and all other entries 0. For $0 \le j \le p, \ 0 \le k \le p+1$, let $N_{j,k}(\underline{x})$ be the $N_j \times N_k$ matrix with its (l, m) element given by

$$\Big[N_{j,k}(\underline{x})\Big]_{l,m} = \sum_{i=1}^{d} D^{\underline{e}_i}\{fg\}(\underline{x}) \int K(\underline{u})\underline{u}^{\tau_j(l)+\tau_k(m)+\underline{e}_i} d\underline{u}, \tag{25}$$

and use these $N_{j,k}(\underline{x})$ to construct a $N \times N$ matrix $N_p(\underline{x})$ and a $N \times N_{p+1}$ matrix $\tilde{M}(\underline{x})$ via

$$N_p(\underline{x}) = \begin{bmatrix} N_{0,0}(\underline{x}) & N_{0,1}(\underline{x}) & \cdots & N_{0,p}(\underline{x}) \\ N_{1,0}(\underline{x}) & N_{1,1}(\underline{x}) & \cdots & N_{1,p}(\underline{x}) \\ \vdots & \ddots & \vdots & \\ N_{p,0}(\underline{x}) & N_{p,1}(\underline{x}) & \cdots & N_{p,p}(\underline{x}) \end{bmatrix}, \quad \tilde{M}(\underline{x}) = \begin{bmatrix} N_{0,p+1}(\underline{x}) \\ N_{1,p+1}(\underline{x}) \\ \vdots \\ N_{p,p+1}(\underline{x}) \end{bmatrix}.$$

Then $S_{n,p}(\underline{x}) = \{fg\}(\underline{x})S_p + hN_p(\underline{x}) + O(h^2)$, $B_{n1}(\underline{x}) = \{fg\}(\underline{x})B_1 + h\tilde{M}(\underline{x}) + O(h^2)$ and $B_{n2}(\underline{x}) = \{fg\}(\underline{x})B_2 + O(h)$. As $S_{n,p}^{-1}(\underline{x}) = \{fg\}^{-1}(\underline{x})S_p^{-1} - h\{fg\}^{-2}(\underline{x})S_p^{-1}N_p(\underline{x})S_p^{-1} + O(h^2)$, we have

$$\begin{aligned} -E\beta_n^*(\underline{x}) =& W_p h^{p+1}\Big[\{fg\}^{-1}(\underline{x})S_p^{-1} - h\{fg\}^{-2}(\underline{x})S_p^{-1}N_p(\underline{x})S_p^{-1}\Big]\Big[\{fg\}(\underline{x})B_1 + h\tilde{M}(\underline{x})\Big]\mathbf{m}_{p+1}(\underline{x}) \\ &+ W_p h^{p+2}\{fg\}^{-1}(\underline{x})S_p^{-1}\{fg\}(\underline{x})B_2\mathbf{m}_{p+2}(\underline{x}) + o(h^{p+2}) \\ =& h^{p+1}W_pS_p^{-1}B_1\mathbf{m}_{p+1}(\underline{x}) + h^{p+2}W_pS_p^{-1}\Big[\{fg\}^{-1}(\underline{x})\mathbf{m}_{p+1}(\underline{x})\{\tilde{M}(\underline{x}) - N_p(\underline{x})S_p^{-1}B_1\} \\ &+ B_2\mathbf{m}_{p+2}(\underline{x})\Big] + o(h^{p+2}). \end{aligned}$$

We claim that for elements $E\beta_{n\underline{r}}^*(\underline{x})$ of $E\beta_n^*(\underline{x})$ with $p - |\underline{r}|$ even, the h^{p+1} term will vanish. This means for any given $\underline{r}$ with $|\underline{r}| \le p$ and $\underline{r}_2$ with $|\underline{r}_2| = p+1$,

$$\sum_{0\le|\underline{r}|\le p} \{S_p^{-1}\}_{N(\underline{r}_1),N(\underline{r})} \ \nu_{\underline{r}+\underline{r}_2} = 0. \tag{26}$$

To prove this, first note that for any $\underline{r}_1$ with $0 \le |\underline{r}_1| \le p$ and $\underline{r}_2$ with $|\underline{r}_2| = p+1$,

$$\sum_{0\le|\underline{r}|\le p} \{S_p^{-1}\}_{N(\underline{r}_1),N(\underline{r})}\ \nu_{\underline{r}+\underline{r}_2} = \int \underline{u}^{\underline{r}_2} K_{\underline{r}_1,p}(\underline{u}) d\underline{u}, \tag{27}$$

where $K_{\underline{r},p}(\underline{u}) = \{|M_{\underline{r},p}(\underline{u})|/|S_p|\}K(\underline{u})$ and $M_{\underline{r},p}(\underline{u})$ is the same as S_p, but with the $N(\underline{r})$ column replaced by $\mu(\underline{u})$. Let c_{ij} denote the cofactor of $\{S_p\}_{i,j}$ and expand the determinant of $M_{\underline{r},p}(\underline{u})$ along the $N(\underline{r})$ column. We see that

$$\int \underline{u}^{\underline{r}_2} K_{\underline{r},p}(\underline{u}) d\underline{u} = |S_p|^{-1} \int \sum_{0\le|\underline{r}|\le p} c_{N(\underline{r}),N(\underline{r}_1)} \underline{u}^{\underline{r}_2+\underline{r}} K(\underline{u}) d\underline{u}.$$

(27) thus follows, because $c_{N(\underline{r}),N(\underline{r}_1)}/|S_p| = \{S_p^{-1}\}_{N(\underline{r}_1),N(\underline{r})}$ from the symmetry of S_p and a standard result concerning cofactors. As a generalization of Lemma 4 in Fan et al (1995) to multivariate case, we can further show that for any $\underline{r}_1$ with $0 \le |\underline{r}_1| \le p$ and $p - |\underline{r}_1|$ even,

$$\int \underline{u}^{\underline{r}_2} K_{\underline{r},p}(\underline{u}) d\underline{u} = 0, \text{ for any } |\underline{r}_2| = p+1,$$

which together with (27) yields to (26). $\square$

We proceed to prove the main results Theorem 3.2. Define $\underline{X}_{ix} = \underline{X}_i - \underline{x}$, $\mu_{ix} = \mu(\underline{X}_{ix})$, $K_{ix} = K_h(\underline{X}_{ix})$ and $\varphi_{ni}(\underline{x};t) = \varphi(Y_i; \mu_{ix}^\top \beta_p(\underline{x}) + t)$. For $\alpha,\ \beta \in \mathcal{R}^N$, define

$$\begin{aligned} \Phi_{ni}(\underline{x};\alpha,\beta) &= K_{ix}\Big\{\rho(Y_i;\mu_{ix}^\top(\alpha+\beta+\beta_p(\underline{x}))) - \rho(Y_i;\mu_{ix}^\top(\beta+\beta_p(\underline{x}))) - \varphi_i(\underline{x};0)\mu_{ix}^\top\alpha\Big\} \\ &= K_{ix}\int_{\mu_{ix}^\top\beta}^{\mu_{ix}^\top(\alpha+\beta)} \{\varphi_{ni}(\underline{x};t) - \varphi_{ni}(\underline{x};0)\}dt, \end{aligned}$$

and $R_{ni}(\underline{x};\alpha,\beta) = \Phi_{ni}(\underline{x};\alpha,\beta) - E\Phi_{ni}(\underline{x};\alpha,\beta)$.

Lemma 5.1 *Under assumptions $(A1)-(A6)$, we have for all large $M>0$,*

$$\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha\in B_n^{(1)},\\ \beta\in B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(\underline{x};\alpha,\beta)| \le M^{3/2} d_n \text{ almost surely,} \tag{28}$$

where $B_n^{(i)} = \{\beta\in\mathcal{R}^N : |H_n\beta| \le M_n^{(i)}\},\ i=1,2.$

Proof. Since $\mathcal{D}$ is compact, it can be covered by a finite number T_n of cubes $\mathcal{D}_k = \mathcal{D}_{n,k}$ with side length $l_n = O(\mathrm{T}_n^{-1/d}) = O\{h(nh^d/\log n)^{-(1-\lambda_2)/2}\}$ and centers $\underline{x}_k = \underline{x}_{n,k}$. Write

$$\begin{aligned}
\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(\underline{x};\alpha,\beta)| \le & \max_{1\le k\le \mathrm{T_n}} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Phi_{ni}(\underline{x}_k;\alpha,\beta) - E\Phi_{ni}(\underline{x}_k;\alpha,\beta)\Big| \\
& + \max_{1\le k\le \mathrm{T}_n} \sup_{\underline{x}\in\mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{\Phi_{ni}(\underline{x}_k;\alpha,\beta) - \Phi_{ni}(\underline{x};\alpha,\beta)\Big\}\Big| \\
& + \max_{1\le k\le \mathrm{T}_n} \sup_{\underline{x}\in\mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{E\Phi_{ni}(\underline{x}_k;\alpha,\beta) - E\Phi_{ni}(\underline{x};\alpha,\beta)\Big\}\Big| \\
\equiv & Q_1 + Q_2 + Q_3.
\end{aligned}$$

In Lemma 5.2, it is shown that $Q_2 \le M^{3/2} d_n/3$ almost surely and thus $Q_3 \le M^{3/2} d_n/3$.

Now all we need to do is to quantify Q_1. To this end, we partition $B_n^{(i)}$, $i = 1, 2$, into a sequence of disjoint subrectangles $D_1^{(i)}, \cdots, D_{J_1}^{(i)}$ such that

$$|D_{j_1}^{(i)}| = \sup\Big\{|H_n(\alpha - \beta)| : \alpha, \beta \in D_{j_1}^{(i)}\Big\} \le 2M^{-1} M_n^{(i)}/\log n, \quad 1 \le j_1 \le J_1.$$

Obviously $J_1 \le (M \log n)^N$. Choose a point $\alpha_{j_1} \in D_{j_1}^{(1)}$ and $\beta_{k_1} \in D_{k_1}^{(2)}$. Then

$$\begin{aligned}
Q_1 \quad \le \quad & \max_{\substack{1\le k\le \mathrm{T}_n\\ 1\le j_1,k_1\le J_1}} \sup_{\substack{\alpha \in D_{j_1}^{(1)},\\ \beta \in D_{k_1}^{(2)}}} |\sum_{i=1}^n \{R_{ni}(\underline{x}_k;\alpha_{j_1},\beta_{k_1}) - R_{ni}(\underline{x}_k;\alpha,\beta)\}| \\
& + \max_{\substack{1\le k\le \mathrm{T}_n\\ 1\le j_1,k_1\le J_1}} |\sum_{i=1}^n R_{ni}(\underline{x}_k;\alpha_{j_1},\beta_{k_1})| = H_{n1} + H_{n2}. \qquad (29)
\end{aligned}$$

We first consider H_{n1}. For each $j_1 = 1, \cdots, J_1$ and $i = 1, 2$, partition each rectangle $D_{j_1}^{(i)}$ further into a sequence of subrectangles $D_{j_1,1}^{(i)}, \cdots, D_{j_1,J_2}^{(i)}$. Repeat this process recursively as follows. Suppose after the lth round, we get a sequence of rectangles $D_{j_1,j_2,\cdots,j_l}^{(i)}$ with $1 \le j_k \le J_k$, $1 \le k \le l$, then in the $(l+1)$th round, each rectangle $D_{j_1,j_2,\cdots,j_l}^{(i)}$ is partitioned into a sequence of subrectangles $\{D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}, 1 \le j_l \le J_l\}$ such that

$$|D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}| = \sup\Big\{|H_n(\alpha - \beta)| : \alpha, \beta \in D_{j_1,j_2,\cdots,j_l,j_{l+1}}^{(i)}\Big\} \le 2M_n^{(i)}/(M^l \log n), \; 1 \le j_{l+1} \le J_{l+1},$$

where $J_{l+1} \le M^N$. End this process after the (L_n+1)th round, with L_n given at the beginning of Section 3. Let $D_l^{(i)},\ i=1,2$, denote the set of all subrectangles of $D_0^{(i)}$ after the lth round of partition and a typical element $D_{j_1,j_2,\cdots,j_l}^{(i)}$ of $D_l^{(i)}$ is denoted as $D_{(j_l)}^{(i)}$. Choose a point $\alpha_{(j_l)} \in D_{(j_l)}^{(1)}$ and $\beta_{(j_l)} \in D_{(j_l)}^{(2)}$ and define

$$V_l = \sum_{\substack{(j_l),\\(k_l)}} P\Big\{\Big|\sum_{i=1}^{n}\{R_{ni}(\underline{x}_k;\alpha_{j_l},\beta_{k_l}) - R_{ni}(\underline{x}_k;\alpha_{j_{l+1}},\beta_{k_{l+1}})\}\Big| \ge \frac{M^{3/2}d_n}{2^l}\Big\},\ 1 \le l \le \mathrm{L}_n,$$

$$Q_l = \sum_{\substack{(j_l),\\(k_l)}} P\Big\{\sup_{\substack{\alpha\,\in\, D^{(1)}_{(j_l)},\\ \beta\,\in\, D^{(2)}_{(k_l)}}}\Big|\sum_{i=1}^{n}\{R_{ni}(\underline{x}_k;\alpha_{j_l},\beta_{k_l}) - R_{ni}(\underline{x}_k;\alpha,\beta)\}\Big| \ge \frac{M^{3/2}d_n}{2^l}\Big\},\ 1 \le l \le \mathrm{L}_n+1.$$

By (A4), it is easy to see that for any $\alpha \in D^{(1)}_{(j_{\mathrm{L}_n+1})} \in D^{(1)}_{\mathrm{L}_n+1}$ and $\beta \in D^{(2)}_{(k_{\mathrm{L}_n+1})} \in D^{(2)}_{\mathrm{L}_n+1}$,

$$|R_{ni}(\underline{x}_k;\alpha,\beta) - R_{ni}(\underline{x}_k;\alpha_{j_{\mathrm{L}_n+1}},\beta_{k_{\mathrm{L}_n+1}})| \le \frac{CM_n^{(2)}}{M^{\mathrm{L}_n+1}\log n},$$

which together with the choice of L_n implies that $Q_{\mathrm{L}_n+1}=0$. As $Q_l \le V_l + Q_l,\ 1 \le l \le \mathrm{L}_n$,

$$P(H_{n1} > \frac{M^{3/2}d_n}{2}) \le \mathrm{T}_n Q_1 \le \mathrm{T}_n \sum_{l=1}^{\mathrm{L}_n} V_l. \tag{30}$$

To quantify V_l, let

$$W_n = \sum_{i=1}^{n} Z_{ni},\ \ Z_{ni} \equiv R_{ni}(\underline{x}_k;\alpha_{j_l},\beta_{k_l}) - R_{ni}(\underline{x}_k;\alpha_{j_{l+1}},\beta_{j_{l+1}}). \tag{31}$$

Note that by (A2), we have, uniformly in $\underline{x},\ \alpha$ and β, that

$$|\Phi_{ni}(\underline{x};\alpha,\beta)| \le CM_n^{(1)}. \tag{32}$$

Therefore, $|Z_{ni}| \le CM_n^{(1)}$. With Lemma 5.6, we can apply Lemma 5.4 to V_l with

$$B_1 = C_1M_n^{(1)},\ B_2 = nh^d(M_n^{(1)})^2M_n^{(2)}\{M^l\log n\}^{-2/\nu_2},$$

$$r_n = r_n^l \equiv (2^{\nu_2/2}/M)^{2l/\nu_2}r(n),\ q = n/r_n^l,\ \eta = M^{3/2}d_n/2^l,$$

$$\lambda_n = (2C_1M_n^{(1)}r_n^l)^{-1},\ \Psi(n) = Cq^{3/2}/\eta^{1/2}\gamma[r_n^l]\{r_n^lM_n^{(1)}\}^{1/2}.$$

Note that $nM_n^{(1)}/\eta \to \infty$, $r_n^l \to \infty$ for all $1 \le l \le \mathrm{L}_n$ from (19) and

$$\lambda\eta = CM^{1/2}\log nM^{2l/\nu_2}/2^{2l},\ \lambda^2B_2 = C\log n^{1-2/\nu_2}M^{2l/\nu_2}/2^{2l} = o(\lambda\eta),$$

which hold uniformly for all $1 \le l \le \mathrm{L}_n$. Therefore,

$$V_l \le \Big(\prod_{j=1}^{l+1} J_j^2\Big) 4\exp\{-C_1 \log n (M/2^{\nu_2})^{2l/\nu_2}\} + C_2 \tau_n^l,$$

where, as $J_1 \le 2(M\log n)^N$ and $J_l \le 2M^N$ for $2 \le l \le L_n$, τ_n^l is given by

$$\tau_n^l = 4^l M^{2N(l+1)} (\log n)^{2N} n^{3/2} \frac{\gamma[r_n^l]\{M_n^{(1)}\}^{1/2}}{r_n^l \{d_n\}^{1/2}}.$$

It is tedious but easy to check that for M large enough,

$$\mathrm{T}_n \sum_{l=1}^{\mathrm{L}_n} \Big[\Big(\prod_{j=1}^{l+1} J_j^2\Big) 4\exp\{-C_1 \log n (M/2^{\nu_2})^{2l/\nu_2}\}\Big] \text{ is summable over } n. \tag{33}$$

As $\gamma[r_n^l]/r_n^l$ is increasing in l, we have

$$\mathrm{T}_n \sum_{l=1}^{\mathrm{L}_n} \tau_n^l \le \mathrm{T}_n (\log n)^{2N} n^{3/2} \frac{\{M_n^{(1)}\}^{1/2}}{\{d_n\}^{1/2}} \frac{\gamma[r_n^{\mathrm{L}_n}]}{r_n^{\mathrm{L}_n}} \prod_{l=1}^{\mathrm{L}_n} 4^l M^{2N(l+1)},$$

which is again summable over n according to (21). This along with (30) and (33) implies that $H_{n1} \le M^{3/2} d_n/2$ almost surely, by the Borel-Cantelli lemma.

For H_{n2}, first note that

$$P(H_{n2} > \eta) \quad \le \quad \mathrm{T}_n J_1^2 P(|\sum_{i=1}^n R_{ni}(\underline{x};\alpha_{j_1},\beta_{k_1})| > \eta). \tag{34}$$

We apply Lemma 5.4 to quantify $P(|\sum_{i=1}^n R_{ni}(\underline{x};\alpha_{j_1},\beta_{k_1}| > \eta)$, with $r_n = r(n)$, $B_1 = 2C_1 M_n^{(1)}$, $B_2 = C_2 n h^d (M_n^{(1)})^2 M_n^{(2)}$, $\lambda_n = \{r(n) M_n^{(1)}\}^{-1}/4C_1$ and $\eta = M^{3/2} d_n$. Then $nB_1/\eta \to \infty$ and

$$\lambda_n \eta/4 = (nh^d)^{(1-\lambda_2)/2} (\log n)^{(1+\lambda_2)/2} / \{16 C_1 r(n)\} = M^{1/2} \log n/(16C_1),$$

$$\lambda_n^2 B_2 = M^{1/4} (nh^d)^{1-\lambda_2} (\log n)^{\lambda_2} / \{16 C_1^2 r^2(n)\} = M^{1/4} \log n/(16C_1^2),$$

$$\Psi(n) \equiv q_n \{nB_1/\eta\}^{1/2} \gamma[r_n] = \mathrm{T}_n J_1^2 q(n)^{3/2}/\eta^{1/2} \gamma[r(n)] \{r(n) M_n^{(1)}\}^{1/2},$$

where $\Psi(n)$ is summable over n by condition (21). Therefore,

$$P(H_{n2} > \eta) \le 2\mathrm{T}_n J_1^2/n^b + \Psi(n),\ b = \frac{1}{16C_1}(M^{1/2} - M^{1/4}\frac{C_2}{C_1}). \tag{35}$$

By selecting M large enough, we can ensure that (35) is summable. Thus, for M large enough, $H_{n2} \le M^{3/2} d_n$ almost surely. By (57), we know for large M, $Q_1 \le M^{3/2} d_n$ almost surely. $\square$

The quantification of Q_2 is very involved, so we put it as a separate Lemma.

Lemma 5.2 *Under the conditions in Lemma 5.1, $Q_2 \le M^{3/2} d_n/3$ almost surely.*

Proof. Let $\underline{X}_{ik} = \underline{X}_i - \underline{x}_k$, $\mu_{ik} = \mu(\underline{X}_{ik})$ and $K_{ik} = K_h(\underline{X}_{ik})$. It is easy to see that we can write $\Phi_{ni}(\underline{x}_k; \alpha, \beta) - \Phi_{ni}(x; \alpha, \beta) = \xi_{i1} + \xi_{i2} + \xi_{i3}$, where

$$\xi_{i1} = \Big(K_{ik}\mu_{ik} - K_{ix}\mu_{ix}\Big)^\top \alpha \int_0^1 \Big\{\varphi_{ni}(\underline{x}_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(\underline{x}_k; 0)\Big\}\, dt,$$
$$\xi_{i2} = K_{ix}\mu_{ix}^\top \alpha \int_0^1 \Big\{\varphi_{ni}(\underline{x}_k; \mu_{ik}^\top(\beta + \alpha t)) - \varphi_{ni}(x; \mu_{ix}^\top(\beta + \alpha t))\Big\}\, dt,$$
$$\xi_{i3} = K_{ix}\mu_{ix}^\top \alpha \{\varphi_{ni}(x; 0) - \varphi_{ni}(\underline{x}_k; 0)\}.$$

Then $P(Q_2 > M^{3/2} d_n/3) \le \mathrm{T}_n(P_{n1} + P_{n2} + P_{n3})$, where

$$P_{nj} \equiv \max_{1 \le k \le \mathrm{T}_n} P\Big(\sup_{\underline{x} \in \mathcal{D}_k} \sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} |\sum_{i=1}^n \xi_{ij}| \ge M^{3/2} d_n/9\Big),\ j = 1, 2, 3.$$

Based on Borel-Cantelli lemma, $Q_2 \le M^{3/2} d_n$ almost surely, if $\sum_n \mathrm{T}_n P_{nj} < \infty,\ j = 1, 2, 3$.

We first tudy P_{n1}. For any fixed $\alpha \in B_n^{(1)}$ and $\beta \in B_n^{(2)}$, let $I_{ik}^{\alpha,\beta} = 1$, if there exists some $t \in [0, 1]$, such that there are discontinuity points of $\varphi(Y_i; \theta)$ between $\mu_{ik}^\top(\beta_p(\underline{x}_k) + \beta + \alpha t))$ and $\mu_{ik}^\top \beta_p(\underline{x}_k)$; and $I_{ik}^{\alpha,\beta} = 0$, otherwise. Write $\xi_{i1} = \xi_{i1} I_{ik}^{\alpha,\beta} + \xi_{i1}(1 - I_{ik}^{\alpha,\beta})$. Note that by (A3), $|(K_{ik}\mu_{ik} - K_{ix}\mu_{ix})^\top \alpha| \le C_2 M_n^{(1)} l_n/h$. Then by (A2) and the fact that $|\mu_{ik}^\top(\beta + \alpha t)| \le C M_n^{(2)}$, we have $|\xi_{i1}(1 - I_{ik}^{\alpha,\beta})| \le C M_n^{(2)} M_n^{(1)} l_n/h$ uniformly in i, α, β and $\underline{x} \in \mathcal{D}_k$. Define $U_{ik} = I\{|\underline{X}_{ik}| \le 2h\}$, whence $\xi_{i1} = \xi_{i1} U_{ik}$ since $l_n = o(h)$. Therefore,

$$\begin{aligned} P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big|\sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta})\Big| > \frac{M^{3/2} d_n}{18}\Big) &\le P\Big(\sum_{i=1}^n U_{ik} > \frac{M^{1/4} n h^d}{18C}\Big) \\ &\le P\Big(|\sum_{i=1}^n U_{ik} - EU_{ik}| > \frac{M^{1/4} n h^d}{36C}\Big), \end{aligned} \tag{36}$$

where the second inequality follows from the fact that $\mathrm{Var}(\sum_{i=1}^n I\{|\underline{X}_{ik}| \le 2h) = O(nh^d)$ implied by Lemma 5.5. To quantify (36), we apply Lemma 5.4 with $B_1 = 1,\ \eta = M^{1/4} n h^d/(18C),\ B_2 = nh^d,\ r_n = r(n)$. As $\lambda_n \eta = C M^{1/4} \log n (nh^d/\log n)^{(1+\lambda_2)/2}$, $\lambda_n^2 B_2 = o(\lambda_n \eta)$ and $\mathrm{T}_n \Psi_n$ is summable over n under condition (21), we know that

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \Big|\sum_{i=1}^n \xi_{i1}(1 - I_{ik}^{\alpha,\beta})\Big| > M^{3/2} d_n/18\Big) \text{ is summable over } n, \tag{37}$$

whence $\sum_n \mathrm{T}_n P_{n1} < \infty$, is equivalent to

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \xi_{i1} I_{ik}^{\alpha,\beta} \Big| > M^{3/2} d_n/18 \Big) \text{ is summable over } n. \tag{38}$$

To prove (38), first note that $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in S_{i;k}^{\alpha,\beta}\}$, where

$$\begin{aligned}
S_{i;k}^{\alpha,\beta} &= \bigcup_{j=1}^m \bigcup_{t\in[0,1]} [a_j - A(\underline{X}_i, \underline{x}_k) + \mu_{ik}^\top(\beta+\alpha t), a_j - A(\underline{X}_i, \underline{x}_k)] \\
&\subseteq \bigcup_{j=1}^m [a_j - CM_n^{(2)}, a_j + CM_n^{(2)}] \equiv D_n, \quad \text{for some } C>0, \\
A(\underline{x}_1, \underline{x}_2) &= (p+1) \sum_{|\underline{r}|=p+1} \frac{1}{\underline{r}!} (\underline{x}_1 - \underline{x}_2)^{\underline{r}} \int_0^1 D^{\underline{r}} m(\underline{x}_2 + w(\underline{x}_1 - \underline{x}_2))(1-w)^p dw,
\end{aligned}$$

where in the derivation of $S_{i;k}^{\alpha,\beta} \subseteq D_n$, we have used the fact that $|\underline{X}_{ik}| \le 2h$ and $A(\underline{X}_i, \underline{x}_k) = O(h^{p+1}) = O(M_n^{(2)})$ uniformly in i. As $I_{ik}^{\alpha,\beta} \le I\{\varepsilon_i \in D_n\}$, we have $|\xi_{i1}| I_{ik}^{\alpha,\beta} \le |\xi_{i1}| U_{ni}$, where $U_{ni} \equiv I(|\underline{X}_{ik}| \le 2h) I\{\varepsilon_i \in D_n\}$, which is independent of the choice of α and β. Therefore,

$$\begin{aligned}
P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \xi_{i1} I_{ik}^{\alpha,\beta} \Big| > M^{3/2} d_n/18 \Big) &\le P\Big(\sum_{i=1}^n U_{ni} > M^{1/2} n h^d M_n^{(2)}/(18C) \Big) \\
&\le P\Big(\sum_{i=1}^n (U_{ni} - EU_{ni}) > \frac{M^{1/2} n h^d M_n^{(2)}}{36C} \Big),
\end{aligned} \tag{39}$$

where the first inequality is because $|\xi_{i1}| \le CM_n^{(1)} l_n/h$ and the second one is because $EU_{ni} = O(h^d M_n^{(2)})$ by (A1). As $EU_{ni}^2 = EU_{ni}$, by Lemma 5.5, we know that $\mathrm{Var}(\sum_{i=1}^n U_{ni}) = Cnh^d M_n^{(2)}$. We can then apply Lemma 5.4 to the last term in (39) with

$$B_2 = Cnh^d M_n^{(2)},\ B_1 \equiv 1,\ r_n = r(n),\ \eta \equiv M^{1/2} n h^d M_n^{(2)}/(36C).$$

Apparently, $\lambda_n \eta = C \log n (nh^d/\log n)^{(1-\lambda_2)/2}$ and $\lambda_n^2 B_2 = o(\lambda_n \eta)$. As in this case $\mathrm{T}_n \Psi_n$ is still summable over n by (21), (38) thus follows.

For P_{n2}, first note that using approach for P_{n1}, we can show that

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big| \sum_{i=1}^n \{\xi_{i2} - \tilde{\xi}_{i2}\} \Big| \ge M^{3/2} d_n/18 \Big) \text{ is summable over } n.$$

where

$$\tilde{\xi}_{i2} = K_{ik}\mu_{ik}^\top \alpha \int_0^1 \Big\{ \varphi_{ni}(\underline{x}_k; \mu_{ik}^\top(\beta+\alpha t)) - \varphi_{ni}(x; \mu_{ix}^\top(\beta+\alpha t)) \Big\}\, dt.$$

Therefore, we would have $\sum \mathrm{T}_n P_{n2} < \infty$, if

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big| \sum_{i=1}^n \tilde{\xi}_{i2} \Big| \geq M^{3/2} d_n/18 \Big) \text{ is summable over } n. \tag{40}$$

For any fixed $\alpha \in B_n^{(1)}$, $\beta \in B_n^{(2)}$ and $\underline{x} \in \mathcal{D}_k$, let $I_{i;k,x}^{\alpha,\beta} = 1$, if there exists some interval $[t_1, t_2] \subseteq [0, 1]$, such that

$$Y_i - \mu_{ik}^\top(\beta_p(\underline{x}_k) + \beta + \alpha t) \leq a_j \leq Y_i - \mu_{ix}^\top(\beta_p(\underline{x}) + \beta + \alpha t),\ \forall t \in [t_1, t_2] \tag{41}$$

with $a_j \in \{a_1, \cdots, a_m\}$; and $I_{i;k,x}^{\alpha,\beta} = 0$, otherwise. Write $\tilde{\xi}_{i2} = \tilde{\xi}_{i2} I_{i;k,x}^{\alpha,\beta} + \tilde{\xi}_{i2}(1 - I_{i;k,x}^{\alpha,\beta})$. Note that $K_{ik}\mu_{ik}^\top\alpha = O(M_n^{(1)})$ and $\varphi_{ni}(\underline{x}_k; \mu_{ik}^\top(\beta+\alpha t)) - \varphi_{ni}(x; \mu_{ix}^\top(\beta+\alpha t)) = O(M_n^{(2)} l_n/h)$ if $I_{i;k,x}^{\alpha,\beta} = 0$. Then again as $\tilde{\xi}_{i2} = \tilde{\xi}_{i2} I\{|\underline{X}_{ik}| \leq 2h\}$, we have similar to (37) that

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \Big| \sum_{i=1}^n \tilde{\xi}_{i2}(1 - I_{i;k,x}^{\alpha,\beta}) \Big| > M^{3/2} d_n/18 \Big) \text{ is summable over } n.$$

Therefore, by (40), to show $\sum \mathrm{T}_n P_{n2} < \infty$, it is sufficient to show that

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big| \sum_{i=1}^n \tilde{\xi}_{i2} I_{i;k,x}^{\alpha,\beta} \Big| \geq M^{3/2} d_n/36 \Big) \text{ is summable over } n. \tag{42}$$

To this end, define $\epsilon_i = \varepsilon_i + A(\underline{X}_i, \underline{x}_k)$. Then $I_{i;k,x}^{\alpha,\beta} = 1$, i.e. (41) is equivalent to

$$A(\underline{X}_i, \underline{x}_k) - A(\underline{X}_i, \underline{x}) + \mu_{ix}^\top(\beta + \alpha t) \leq \epsilon_i - a_j \leq \mu_{ik}^\top(\beta + \alpha t),\ \forall t \in [t_1, t_2]. \tag{43}$$

Let $\delta_n \equiv M_n^{(2)} l_n/h$. Then $|A(\underline{X}_i, \underline{x}_k) - A(\underline{X}_i, \underline{x})| \leq C\delta_n$, $|(\mu_{ik} - \mu_{ix})^\top \beta| \leq C\delta_n$ and (43) thus implies that

$$-2C\delta_n + \mu_{ik}^\top(\beta + \alpha t) \leq \epsilon_i - a_j \leq \mu_{ik}^\top(\beta + \alpha t) + 2C\delta_n, \quad \forall t \in [t_1, t_2]. \tag{44}$$

Without loss of generality, assume $\mu_{ik}^\top \alpha > 0$. Then from (44) we can see that

$$-2C\delta_n + \mu_{ik}^\top(\beta + \alpha t_2) \leq \epsilon_i - a_j \leq \mu_{ik}^\top(\beta + \alpha t_1) + 2C\delta_n, \tag{45}$$

which in turn means that if $I^{\alpha,\beta}_{i;k,x} = 1$, then $|\xi_{i2}| \leq C(t_2 - t_1)|\mu_{ik}^\top \alpha| \leq 4C\delta_n$ uniformly in i, $\alpha \in B_n^{(1)}$, $\beta \in B_n^{(2)}$ and $\underline{x} \in \mathcal{D}_k$. Therefore, as $\tilde{\xi}_{i2} = \tilde{\xi}_{i2} I\{|\underline{X}_{ik}| \leq 2h\}$, we have

$$
\begin{aligned}
& P\Big(\sup_{\substack{\alpha \in B_n^{(1)} \\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big| \sum_{i=1}^n \tilde{\xi}_{i2} I^{\alpha,\beta}_{i;k,x} \Big| \geq \frac{M^{3/2} d_n}{36} \Big) \\
& \leq P\Big(\sup_{\substack{\alpha \in B_n^{(1)} \\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} I^{\alpha,\beta}_{i;k,x} \geq \frac{M^{5/4} nh^d M_n^{(1)}}{36C} \Big). \qquad (46)
\end{aligned}
$$

We will bound $I^{\alpha,\beta}_{i;k,x}$ by a random variable that is independent of the choice of $\alpha \in B_n^{(1)}$ and $\underline{x} \in D_k$. By the definition of $I^{\alpha,\beta}_{i;k,x}$ and (45), the necessary condition for $I^{\alpha,\beta}_{i;k,x} = 1$ is

$$
\epsilon_i \in \bigcup_{j=1}^m [a_j + \mu_{ik}^\top \beta - 2M_n^{(1)}, a_j + \mu_{ik}^\top \beta + 2M_n^{(1)}] \equiv D^\beta_{ni}, \qquad (47)
$$

which is indeed independent of the choice of α and $\underline{x} \in \mathcal{D}_k$. Therefore,

$$
\begin{aligned}
& P\Big(\sup_{\substack{\alpha \in B_n^{(1)}, \\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} I^{\alpha,\beta}_{i;k,x} \geq \frac{M^{5/4} nh^d M_n^{(1)}}{36C} \Big) \\
\leq \quad & P\Big(\sup_{\beta \in B_n^{(2)}} \sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} I\{\epsilon_i \in D^\beta_{ni}\} \geq \frac{M^{5/4} nh^d M_n^{(1)}}{36C} \Big). \qquad (48)
\end{aligned}
$$

Now we partition $B_n^{(2)}$ into a sequence of subrectangles $S_1, \cdots, S_m$, such that

$$
|S_l| = \sup\Big\{ |H_n(\beta - \beta')| : \beta, \beta' \in S_l \Big\} \leq M_n^{(1)}, \quad 1 \leq l \leq m.
$$

Obviously, $m \leq (M_n^{(2)}/M_n^{(1)})^N = M^{-3N/4}(nh^d/\log n)^{(\lambda_1 - \lambda_2)N}$. Choose a point $\beta_l \in S_l$ for each $1 \leq l \leq m$, and thus

$$
\begin{aligned}
& P\Big(\sup_{\beta \in B_n^{(2)}} \sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} I\{\epsilon_i \in D^\beta_{ni}\} \geq \frac{M^{5/4} nh^d M_n^{(1)}}{36C} \Big) \\
\leq \quad & mP\Big(\sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} I\{\epsilon_i \in D^{\beta_l}_{ni}\} \geq \frac{M^{5/4} nh^d M_n^{(1)}}{72C} \Big) \\
& + mP\Big(\sup_{\beta' \in S_l} \sum_{i=1}^n I\{|\underline{X}_{ik}| \leq 2h\} |I\{\epsilon_i \in D^{\beta_l}_{ni}\} - I\{\epsilon_i \in D^{\beta'}_{ni}\}| \geq \frac{M^{5/4} nh^d M_n^{(1)}}{72C} \Big) \\
\equiv \quad & m(T_1 + T_2). \qquad (49)
\end{aligned}
$$

We deal with T_1 first. Let

$$U^j_{ni} \equiv I\{|\underline{X}_{ik}| \le 2h\} I\{\epsilon_i \in D^{\beta_l}_{ni}\}. \tag{50}$$

Then by the definition of $D^{\beta_j}_{ni}$ given in (47), $EU^j_{ni} = O(h^d M^{(1)}_n) < M^{5/4} h^d M^{(1)}_n/(144C)$ for large M and we have

$$T_1 \le P\Big(\sum_{i=1}^n (U^j_{ni} - EU^j_{ni}) \ge \frac{M^{5/4} n h^d M^{(1)}_n}{144C}\Big).$$

We can thus apply Lemma 5.4 to the quantity on the right hand side with $B_1 \equiv 1$, B_2 given by (69), $r_n = r(n)$ and $\eta \propto M^{5/4} n h^d M^{(1)}_n$, and $\lambda_n = 1/(2r_n)$. It follows that

$$\lambda_n \eta = C M^{5/4} \log n (nh^d/\log n)^{(1+\lambda_2)/2 - \lambda_1},\ \lambda_n^2 B_2 = C \log n (nh^d/\log n)^{-2(\lambda_1-\lambda_2)/\nu_2}.$$

As $(1+\lambda_2)/2 \ge \lambda_1$ and $\lambda_2 < \lambda_1$, we have $T_1 = O(n^{-b})$ for any $b > 0$.

For T_2, note that as $|\mu^\top_{ik}(\beta - \beta_l)| \le CM^{(1)}_n$ for any $\beta \in S_l$, $1 \le l \le m$, we have

$$\begin{aligned} |I\{\epsilon_i \in D^{\beta_l}_{ni}\} - I\{\epsilon_i \in D^{\beta}_{ni}\}| &= I\{\epsilon_i \in D^{\beta_l}_{ni} \setminus D^{\beta}_{ni}\} \\ &\le I\Big\{\epsilon_i \in \bigcup_{j=1}^m [a_j + \mu^\top_{ik}\beta_l - CM^{(1)}_n, a_j + \mu^\top_{ik}\beta_l + CM^{(1)}_n]\Big\} \equiv U_{ni}, \end{aligned}$$

for some $C > 0$, which is independent of the choice of $\beta \in S_l$. Therefore,

$$T_2 \le P\Big(\sum_{i=1}^n I\{|\underline{X}_{ik}| \le 2h\} U_{ni} \ge \frac{M^{5/4} n h^d M^{(1)}_n}{72C}\Big),$$

which can be dealt with similarly as with T_1 and thus $T_2 = O(n^{-b})$ for any $b > 0$. Thus from (46), (48) and (49), we can claim that (42) is true and thus $\mathrm{T}_n P_{n2}$ is summable over n.

The quantification of P_{n3} is much simpler, as no β is involved in ξ_{i3}. For any given $\underline{x} \in \mathcal{D}_k$, let $I_{i;k,x} = 1$, if there is a discontinuity point of $\varphi(Y_i;\theta)$ between $\mu^\top_{ik}\beta_p(\underline{x}_k)$ and $\mu^\top_{ix}\beta_p(\underline{x})$; and $I_{i;k,x} = 0$ otherwise. Write $\xi_{i3} = \xi_{i3} I_{i;k,x} + \xi_{i3}(1 - I_{i;k,x})$. Again by (A2) and the fact that $|K_{ix}\mu^\top_{ix}\alpha| = O(M^{(1)}_n)$ and $|\mu^\top_{ik}\beta_p(\underline{x}_k) - \mu^\top_{ix}\beta_p(\underline{x})| = |A(\underline{X}_i, \underline{x}_k) - A(\underline{X}_i, \underline{x})| = O(M^{(2)}_n l_n/h)$, we have similar to (37) that

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha \in B^{(1)}_n \\ \underline{x} \in \mathcal{D}_k}} \Big|\sum_{i=1}^n \xi_{i3}(1 - I_{i;k,x})\Big| > M^{3/2} d_n/18\Big) \text{ is summable over } n.$$

It's easy to see that $I_{i;k,x} \le I\{\varepsilon_i + A(\underline{X}_i, \underline{x}_k) \in S_{i;k,x}\}$, where

$$\begin{aligned} S_{i;k,x} &= \bigcup_{j=1}^{m} \bigcup_{t\in[0,1]} \Big[a_j - |A(\underline{X}_i, \underline{x}_k) - A(\underline{X}_i, \underline{x})|, a_j + |A(\underline{X}_i, \underline{x}_k) - A(\underline{X}_i, \underline{x})|\Big] \\ &\subseteq \bigcup_{j=1}^{m} [a_j - CM_n^{(2)} l_n/h, a_j + CM_n^{(2)} l_n/h] \equiv D_n, \text{ for some } C > 0. \end{aligned}$$

Therefore, $|\xi_{i3}| I_{i;k,x} = |\xi_{i3}| I\{|\underline{X}_{ik}| \le 2h\} I_{i;k,x} \le U_{ni}$, where

$$U_{ni} \equiv M_n^{(1)} I\{|\underline{X}_{ik}| \le 2h\} I\{\varepsilon_i + A(\underline{X}_i, \underline{x}_k) \in D_n\},$$

which is independent of the choice of $\alpha \in B_n^{(1)}$ and $\underline{x} \in \mathcal{D}_k$. Thus

$$\mathrm{T}_n P\Big(\sup_{\substack{\alpha\,\in\, B_n^{(1)} \\ \underline{x}\,\in\,\mathcal{D}_k}} \Big|\sum_{i=1}^{n} \xi_{i3} I_{i;k,x}\Big| > M^{3/2} d_n/18\Big) \le \mathrm{T}_n P\Big(\sum_{i=1}^{n} [U_{ni} - EU_{ni}] > M^{3/2} d_n/36\Big), \tag{51}$$

where we have used the fact that $EU_{ni} = O(h^d M_n^{(1)} M_n^{(2)} l_n/h) = O(d_n/n)$. We will have $\sum \mathrm{T}_n P_{n3} < \infty$ if the right hand side in (51) is summable over n, i.e.

$$\mathrm{T}_n P\Big(\sum_{i=1}^{n} [U_{ni} - EU_{ni}] > M^{3/2} d_n/36\Big) \text{ is summable over } n. \tag{52}$$

It's easy to check that Lemma 5.5 again holds with $\psi_{\underline{x}}(\underline{X}_i, Y_i)$ standing for U_{ni}. Applying Lemma 5.4 to (52) with $B_1 \equiv M_n^{(1)}$, $B_2 \equiv Cnh^d (M_n^{(1)})^2 M_n^{(2)} l_n/h$, $\eta \equiv M^{3/2} d_n/36$ and $r_n = r(n)$, we have (note that $nB_1/\eta \to \infty$ indeed)

$$\lambda_n \eta/4 = CM^{1/2} \log n, \;\; \lambda_n^2 B_2 = C r_n^{-2/\nu_2} \log n = o(\lambda_n \eta).$$

Thus, $\mathrm{T}_n \Psi_n$ again is summable over n and (52) indeed holds. $\square$

Proof of Theorem 3.2. Let $\lambda_1 = \lambda(s)$. Then according to Lemma 5.1 and Lemma 5.9, we know that with probability 1, there exists some $C_1 > 1$, such that for all large $M > 0$,

$$\begin{aligned} &\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha\,\in\, B_n^{(1)}, \\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^{n} \Phi_{ni}(\underline{x}; \alpha, \beta) - \frac{nh^d}{2} (H_n \alpha)^\top S_{np}(\underline{x}) H_n(\alpha + 2\beta)\Big| \\ &\le C_1 M^{3/2} (d_{n1} + d_n) \le 2C_1 M^{3/2} (nh^d)^{1-2\lambda_1} (\log n)^{2\lambda_1}, \text{ when } n \text{ is large,} \end{aligned} \tag{53}$$

where $d_{n1} = (nh^d)^{1-\lambda_1-2\lambda_2} (\log n)^{\lambda_1+2\lambda_2}$. Note that from (12), we can write

$$\sum_{i=1}^{n} K_{ni} \varphi(Y_i; \mu_{ni}^\top \beta_p(\underline{x})) \mu_{ni}^\top \alpha = nh^d \beta_n^*(\underline{x})^\top W_p^{-1} S_{np}(\underline{x}) H_n \alpha.$$

Replace $B_n^{(1)}$ in (53) with $B_{nk}^{(1)} = \left\{\alpha \in \mathcal{R}^N : k \leq M^{-1}(nh^d/\log n)^{\lambda_1}|H_n\alpha| \leq k+1\right\}$ and M with $(k+1)M$. We have, by the definition of $\Phi_{ni}(\underline{x};\alpha,\beta)$, that

$$
\begin{aligned}
& \inf_{\underline{x}\in\mathcal{D}} \inf_{\substack{\alpha \in B_{nk}^{(1)}, \\ \beta \in B_n^{(2)}}} \Big\{ \sum_{i=1}^n \rho(Y_i; \mu_{ni}^\top(\alpha+\beta+\beta_p(\underline{x})))K_{ni} - \sum_{i=1}^n \rho(Y_i; \mu_{ni}^\top(\beta+\beta_p(\underline{x})))K_{ni} \\
& \qquad + nh^d(W_p^{-1}\beta_n^*(\underline{x}) - H_n\beta)^\top S_{np}(\underline{x})H_n\alpha \Big\} \\
\geq\; & \inf_{\underline{x}\in\mathcal{D}} \inf_{\alpha\in B_{nk}^{(1)}} \frac{nh^d}{2}(H_n\alpha)^\top S_{np}(\underline{x})H_n\alpha - 2CM^{3/2}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} \\
\geq\; & \left\{C_3(kM)^2/2 - 2C_1(k+1)^{3/2}M^{3/2}\right\}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} \\
\geq\; & (8-2^{5/2})C_1C_4^{3/2}(nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1} > 0 \text{ almost surely,} \qquad (54)
\end{aligned}
$$

where the last term is independent of the choice of $k \geq 1$. The last inequality is derived as follows. As $S_p > 0$, suppose its minimum eigenvalue is $\tau_1 > 0$. As $S_{np}(\underline{x}) \to g(\underline{x})f(\underline{x})S_p$ uniformly in $\underline{x} \in \mathcal{D}$ by Lemma 5.8 and $g(\underline{x})f(\underline{x})$ is bounded away from zero by (A5) and (24), there exists some constant $C_3 > 0$, such that for all $\underline{x} \in \mathcal{D}$, the minimum eigenvalue of $S_{np}(\underline{x})$ is greater than C_3. The last inequality thus holds if $M \geq C_4 = (16C_1/C_3)^2$. Note that

$$
\bigcup_{k=1}^{\infty} B_{nk}^{(1)} = \left\{\alpha| \in \mathcal{R}^N : \left(\frac{nh^d}{\log n}\right)^{\lambda_1}|H_n\alpha| \geq M\right\} := B_n^N. \qquad (55)
$$

Therefore, from (54) and (55), we have

$$
\begin{aligned}
& \inf_{\underline{x}\in\mathcal{D}} \inf_{\substack{\alpha \in B_n^N, \\ \beta \in B_n^{(2)}}} \Big\{ \sum_{i=1}^n \rho(Y_i; \mu_{ni}^\top(\alpha+\beta+\beta_p(\underline{x})))K_{ni} - \sum_{i=1}^n \rho(Y_i; \mu_{ni}^\top(\beta+\beta_p(\underline{x})))K_{ni} \\
& \qquad + nh^d(W_p^{-1}\beta_n^*(\underline{x}) - H_n\beta)^\top S_{np}(\underline{x})H_n\alpha \Big\} > 0 \text{ almost surely.} \qquad (56)
\end{aligned}
$$

Note that by (58), Lemma 5.10 and Proposition 3.1, we have $|\beta_n^*(\underline{x})| \leq C_3(nh^d/\log n)^{-\lambda_2}$ uniformly in $\underline{x} \in \mathcal{D}$ almost surely. Namely, $\beta_n^*(\underline{x}) \in B_n^{(2)}$ for all $\underline{x} \in \mathcal{D}$, if $M > C_3^4$. This implies that if $M > \max(C_3^4, C_4)$, (56) still holds with β replaced with $H_n^{-1}W_p^{-1}\beta_n^*(\underline{x})$. Therefore,

$$
\begin{aligned}
& \inf_{\underline{x}\in\mathcal{D}} \inf_{\alpha\in B_n^N} \Big\{ \sum_{i=1}^n K_{ni}\rho(Y_i; \mu_{ni}^\top(\alpha + H_n^{-1}W_p^{-1}\beta_n^*(\underline{x}) + \beta_p(\underline{x}))) \\
& \qquad\qquad - \sum_{i=1}^n K_{ni}\rho(Y_i; \mu_{ni}^\top(H_n^{-1}W_p^{-1}\beta_n^*(\underline{x}) + \beta_p(\underline{x}))) \Big\} > 0,
\end{aligned}
$$

which is equivalent to Theorem 3.2. $\square$

Proof of (13). Let $\tilde{d}_n = (nh^d)^{1-2\lambda_1}(\log n)^{2\lambda_1}$. Through the proof lines of Theorem 3.2, we can see that (13) will follow if

$$\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(\underline{x};\alpha,\beta)| \le M^{3/2}\tilde{d}_n \text{ almost surely,}$$

with $\lambda_1 = 1,\ \lambda_2 = 1/2$ and $B_n^{(i)},\ i = 1,2$ defined as in Lemma 5.1.

To prove this, cover $\mathcal{D}$ by a finite number $\tilde{\mathrm{T}}_n = \{(nh^d/\log n)^{1/2}/h\}^d$ of cubes $\mathcal{D}_k = \mathcal{D}_{nk}$ with side length $\tilde{l}_n = O\{h(nh^d/\log n)^{-1/2}\}$ and centers $\underline{x}_k = \underline{x}_{n,k}$. Write

$$\begin{aligned}
\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} |\sum_{i=1}^n R_{ni}(\underline{x};\alpha,\beta)| \le{}& \max_{1\le k\le\tilde{\mathrm{T}}_n} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^n \Phi_{ni}(\underline{x}_k;\alpha,\beta) - E\Phi_{ni}(\underline{x}_k;\alpha,\beta)\Big| \\
&+ \max_{1\le k\le\tilde{\mathrm{T}}_n} \sup_{\underline{x}\in\mathcal{D}_k} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{\Phi_{ni}(\underline{x}_k;\alpha,\beta) - \Phi_{ni}(\underline{x};\alpha,\beta)\Big\}\Big| \\
&+ \max_{1\le k\le\tilde{\mathrm{T}}_n} \sup_{\underline{x}\in\mathcal{D}_k} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} \Big|\sum_{i=1}^n \Big\{E\Phi_{ni}(\underline{x}_k;\alpha,\beta) - E\Phi_{ni}(\underline{x};\alpha,\beta)\Big\}\Big| \\
\equiv{}& Q_1 + Q_2 + Q_3.
\end{aligned}$$

We will show that with probability 1, $Q_k \le M^{3/2}\tilde{d}_n/3,\ k = 1,2,3$.

Define ξ_{ij} as in Lemma 5.1. As $P(Q_2 > M^{3/2}\tilde{d}_n/2) \le \tilde{\mathrm{T}}_n(P_{n1} + P_{n2} + P_{n3})$, where

$$P_{nj} \equiv \max_{1\le k\le\tilde{\mathrm{T}}_n} P\Big(\sup_{\underline{x}\in\mathcal{D}_k} \sup_{\substack{\alpha\,\in\, B_n^{(1)},\\ \beta\,\in\, B_n^{(2)}}} |\sum_{i=1}^n \xi_{ij}| \ge M^{3/2}\tilde{d}_n/9\Big),\ j = 1,2,3.$$

Then based on Borel-Cantelli lemma, $Q_2 \le M^{3/2}\tilde{d}_n/2$ almost surely if $\sum_n \tilde{\mathrm{T}}_n P_{nj} < \infty$, for $j = 1,2,3$. We only prove that for P_{n1} to illustrate. Recall that

$$\xi_{i1} = \Big(K_{ik}\mu_{ik} - K_{ix}\mu_{ix}\Big)^\top \alpha \int_0^1 \Big\{\varphi_{ni}(\underline{x}_k;\mu_{ik}^\top(\beta+\alpha t)) - \varphi_{ni}(\underline{x}_k;0)\Big\}\,dt.$$

Because $|(K_{ik}\mu_{ik} - K_{ix}\mu_{ix})^\top\alpha| \le C_2 M_n^{(1)}\tilde{l}_n/h,\ |\mu_{ik}^\top(\beta+\alpha t)| \le CM_n^{(2)}$ and $\varphi(.)$ is Lipschitz continuous, we have $|\xi_{i1}| \le CM_n^{(2)}M_n^{(1)}\tilde{l}_n/h$. Define $U_{ik} = I\{|\underline{X}_{ik}| \le 2h\}$. As $\tilde{l}_n = o(h)$, we can

see that $\xi_{i1} = \xi_{i1} U_{ik}$ and similar to (36), we have

$$\begin{aligned} P\Big(\sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} \sup_{\underline{x} \in \mathcal{D}_k} \Big|\sum_{i=1}^n \xi_{i1}\Big| > \frac{M^{3/2}\tilde{d}_n}{9}\Big) &\leq P\Big(\sum_{i=1}^n U_{ik} > \frac{M^{1/4} n h^d}{9C}\Big) \\ &\leq P\Big(|\sum_{i=1}^n U_{ik} - EU_{ik}| > \frac{M^{1/4} n h^d}{18C}\Big), \end{aligned}$$

and $\sum_n \tilde{\mathrm{T}}_n P_{nj} < \infty$ thus follows from similar arguments as those lying between (36) and (37). The proof of $Q_1 \leq M^{3/2}\tilde{d}_n/2$ almost surely is much easier than in Lemma 5.1, if $\varphi(.)$ is Lipschitz continuous. Instead of the iterative partition approach adopted there, we once for all partition $B_n^{(i)},\ i = 1, 2,$ into a sequence of disjoint subrectangles $D_1^{(i)}, \cdots, D_{J_1}^{(i)}$ such that

$$|D_{j_1}^{(i)}| = \sup\Big\{|H_n(\alpha - \beta)| : \alpha, \beta \in D_{j_1}^{(i)}\Big\} \leq M_n^{(i)} (\log n/n)^{1/2}, \quad 1 \leq j_1 \leq J_1.$$

Obviously $J_1 \leq (n/\log n)^{N/2}$. Choose a point $\alpha_{j_1} \in D_{j_1}^{(1)}$ and $\beta_{k_1} \in D_{k_1}^{(2)}$. Then

$$\begin{aligned} Q_1 \leq & \max_{\substack{1 \leq k \leq \tilde{\mathrm{T}}_n \\ 1 \leq j_1, k_1 \leq J_1}} \sup_{\substack{\alpha \in D_{j_1}^{(1)},\\ \beta \in D_{k_1}^{(2)}}} |\sum_{i=1}^n \{R_{ni}(\underline{x}_k; \alpha_{j_1}, \beta_{k_1}) - R_{ni}(\underline{x}_k; \alpha, \beta)\}| \\ & + \max_{\substack{1 \leq k \leq \mathrm{T}_n \\ 1 \leq j_1, k_1 \leq J_1}} |\sum_{i=1}^n R_{ni}(\underline{x}_k; \alpha_{j_1}, \beta_{k_1})| = H_{n1} + H_{n2}. \end{aligned} \tag{57}$$

By Lipschitz continuity of $\varphi(.)$, we have for any $\alpha \in D_{j_1}^{(1)}$ and $\beta \in D_{k_1}^{(2)}$,

$$|\Phi_{ni}(\underline{x}_k; \alpha_{j_1}, \beta_{k_1}) - \Phi_{ni}(\underline{x}_k; \alpha, \beta)|^2 = O(\{M_n^{(2)}\}^3 \log n/n) < M^{3/2}\tilde{d}_n/(4n).$$

Therefore, it remains to show that $P(H_{n2} > M^{3/2}\tilde{d}_n/4)$ is summable over n.

First note that by Cauchy inequality, $|R_{ni}(\underline{x}; \alpha, \beta)|^2 = O(\{M_n^{(1)} M_n^{(2)}\}^2)$ and $E|R_{ni}(\underline{x}; \alpha, \beta)|^2 = O(h^d \{M_n^{(1)} M_n^{(2)}\}^2)$ uniformly in $\underline{X}_i,\ \underline{x},\ \alpha \in M_n^{(1)}$ and $\beta \in M_n^{(2)}$. Next, for any $\eta > 0$,

$$P(H_{n2} > \eta) \quad \leq \quad \tilde{\mathrm{T}}_n J_1^2 P(|\sum_{i=1}^n R_{ni}(\underline{x}; \alpha_{j_1}, \beta_{k_1})| > \eta).$$

We apply Lemma 5.4 with $r_n = (nh^d/\log n)^{1/2},\ B_1 = 2C_1 M_n^{(1)} M_n^{(2)},\ B_2 = C_2 n h^d (M_n^{(1)} M_n^{(2)})^2,\ \lambda_n = (4C_1 r_n \{M_n^{(2)}\}^2)^{-1}$ and $\eta = M^{3/2}\tilde{d}_n/4$. It is easy to see that $nB_1/\eta \to \infty$ and

$$\lambda_n \eta/4 = M \log n/(16C_1),\ \lambda_n^2 B_2 = o(\lambda_n \eta)$$

$$\Psi(n) \equiv q_n \{nB_1/\eta\}^{1/2} \gamma[r_n] = n^{3/2} (\log n)^{-1/2} \gamma[r(n)]/r(n).$$

As $\tilde{\mathrm{T}}_n J_1^2 \Psi(n)$ is summable over n by condition (21), so is $P(H_{n2} > M^{3/2}\tilde{d}_n/4)$. □

Proof of Corollary 3.3. As $1+\lambda_2 \geq 2\lambda_1$, it's sufficient to prove that with probability 1,

$$\beta_n^*(\underline{x}) - E\beta_n^*(\underline{x}) - \frac{1}{nh^d} W_p S_{np}^{-1}(\underline{x}) H_n^{-1} \sum_{i=1}^n K_h(\underline{X}_i - \underline{x})\varphi(\varepsilon_i)\mu(\underline{X}_i - \underline{x}) = O\Big\{\Big(\frac{\log n}{nh^d}\Big)^{(1+\lambda_2)/2}\Big\}, \quad (58)$$

uniformly in $\underline{x} \in \mathcal{D}$. As $\varphi(\varepsilon_i) \equiv \varphi(Y_i, m(X_i))$ and $E\varphi(\varepsilon_i) = 0$, the term on the left hand side of (58) stands for

$$W_p S_{n,p}^{-1}(\underline{x}) \frac{1}{nh^d} \sum_{i=1}^n \{Z_{ni}(\underline{x}) - EZ_{ni}(\underline{x})\},$$

where

$$Z_{ni}(\underline{x}) = H_n^{-1} K_h(\underline{X}_i - \underline{x})\mu(\underline{X}_i - \underline{x})\Big\{\varphi(Y_i, \mu(\underline{X}_i - \underline{x})^\top \beta_p(\underline{x})) - \varphi(\varepsilon_i)\Big\}.$$

Next, like what we did in Lemma 5.1, we cover $\mathcal{D}$ with number T_n cubes $\mathcal{D}_k = \mathcal{D}_{n,k}$ with side length $l_n = O(\mathrm{T}_n^{-1/d})$ and centers $\underline{x}_k = \underline{x}_{n,k}$. Write

$$\begin{aligned}
\sup_{\underline{x}\in\mathcal{D}} \Big|\sum_{i=1}^n Z_{ni}(\underline{x}) - EZ_{ni}(\underline{x})\Big| \leq & \max_{1\leq k\leq \mathrm{T}_n} \Big|\sum_{i=1}^n Z_{ni}(\underline{x}_k) - EZ_{ni}(\underline{x}_k)\Big| \\
& + \max_{1\leq k\leq \mathrm{T}_n} \sup_{\underline{x}\in\mathcal{D}_k} \Big|\sum_{i=1}^n Z_{ni}(\underline{x}) - Z_{ni}(\underline{x}_k)\Big| \\
& + \max_{1\leq k\leq \mathrm{T}_n} \sup_{\underline{x}\in\mathcal{D}_k} \Big|\sum_{i=1}^n EZ_{ni}(\underline{x}) - EZ_{ni}(\underline{x}_k)\Big| \\
\equiv & Q_1 + Q_2 + Q_3.
\end{aligned}$$

As $Z_{ni}(\underline{x}) - Z_{ni}(\underline{x}_k) = H_n^{-1} K_h(\underline{X}_i - \underline{x})\mu(\underline{X}_i - \underline{x})\{\varphi_{ni}(\underline{x}; 0) - \varphi_{ni}(\underline{x}_k; 0)\}$, through approaches similar to that for ξ_{i3} in the proof of Lemma 5.2, we can show that

$$Q_2 = O\Big\{\Big(\frac{nh^d}{\log n}\Big)^{(1-\lambda_2)/2} \log n\Big\} \text{ almost surely}$$

and so is Q_3. To bound Q_1, first note that $EZ_{ni}^2(\underline{x}_k) = O(h^{p+1+d})$ uniformly in i and k. As $|Z_{ni}(\underline{x})| \leq C$ for some constant C by (A2), we can see that from Lemma 5.5

$$\sum_{i=1}^n EZ_{ni}^2(\underline{x}_k) + \sum_{i<j} |\mathrm{Cov}(Z_{ni}(\underline{x}_k), Z_{nj}(\underline{x}_k))| \leq C_2 n h^{p+1+d}.$$

Finally by Lemma 5.4 with $B_1 = C_1$, $B_2 \equiv Cnh^{p+1+d}$, $\eta = A_3(nh^d/\log n)^{(1-\lambda_2)/2}\log n$ and $r_n = r(n)$, we have (note that $nB_1/\eta \to \infty$ indeed)

$$\lambda_n\eta = A_3/(2C_1)\log n,\ \lambda_n^2 B_2 = C_2/(4C_1^2)\log n.$$

Therefore,

$$P\Big(\max_{1\le k\le \mathrm{T_n}}\Big|\sum_{i=1}^n Z_{ni}(\underline{x}_k) - EZ_{ni}(\underline{x}_k)\Big| \ge A_3(nh^d/\log n)^{(1-\lambda_2)/2}\log n\Big) \le \mathrm{T}_n/n^a + C\mathrm{T}_n\Psi_n,$$

where $a = A_3/(8C_1) - C_2/(4C_1^2)$. By selecting A_3 large enough, we can ensure that T_n/n^a is summable over n. As $\mathrm{T}_n\Psi_n$ is summable over n from (21), we can conclude that

$$Q_1 = O\Big\{\Big(\frac{nh^d}{\log n}\Big)^{(1-\lambda_2)/2}\log n\Big\} \text{ almost surely.}$$

This together with Lemma 5.8 completes the proof. □

Proof of Corollary 4.1. Through the proof lines for Theorem 3.2 and Corollary 3.3, it's not difficult to see that Corollary 3.3 still holds under the conditions imposed here. Under the additive structure (4), we thus have

$$\begin{aligned}\phi_{n1}(x_1) =&\phi_1(x_1) + \frac{1}{n}\sum_{i=1}^n m_2(\underline{X}_{2i}) - h^{p+1}e_1W_pS_p^{-1}B_1\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{p+1}(x_1,\underline{X}_{2i})\\ &+\frac{1}{n^2h_1h^{d-1}}e_1\sum_{j=1}^n\varphi(\varepsilon_j)\sum_{i=1}^n S_{np}^{-1}(x_1,\underline{X}_{2i})K(X_{1,xj}/h_1,\underline{X}_{2,ij}/h)\mu(X_{1,xj}/h_1,\underline{X}_{2,ij}/h)\\ &+o_p(\{\max(h_1,h)\}^{p+1}) + O_p\{(nh_1h^{d-1}/\log n)^{-3/4}\},\end{aligned}\tag{59}$$

where $X_{1,xj} = X_{1j} - x$, $\underline{X}_{2,ij} = \underline{X}_{2i} - \underline{X}_{2j}$ and e_1 is as in Proposition 3.1. Note that by (17), $(nh_1)^{1/2}(nh_1h^{d-1}/\log n)^{-3/4} \to 0$, the $O_p(.)$ term can thus be safely ignored.

By central limit theorem for strongly mixing processes (Bosq, 1998, Theorem 1.7), we have

$$\frac{1}{n}\sum_{i=1}^n m_2(\underline{X}_{2i}) = O_p(n^{-1/2}),\quad \frac{1}{n}\sum_{i=1}^n \mathbf{m}_{p+1}(x_1,\underline{X}_{2i}) = E\mathbf{m}_{p+1}(x_1,\underline{X}_2) + O_p(n^{-1/2}).$$

As the expectations of all other terms in (59) are 0, the leading term in the asymptotic bias of $\tilde{\phi}_1(x_1) - \phi_1(x_1)$ is thus given by

$$-\{\max(h_1,h)\}^{p+1}e_1W_pS_p^{-1}B_1E\mathbf{m}_{p+1}(x_1,\underline{X}_2).$$

Again through standard arguments in Masry (1996), we can see that

$$\begin{aligned}&\frac{1}{nh^{d-1}}\sum_{i=1}^{n} S_{np}^{-1}(x_1,\underline{X}_{2i})K_h(X_{1,xj},\underline{X}_{2,ij})\mu(X_{1,xj}/h_1,\underline{X}_{2,ij}/h)\\ = \quad & S_{np}^{-1}(x_1,\underline{X}_{2j})f_2(\underline{X}_{2j})\int_{[0,1]^{\otimes d-1}}\{K\mu\}(X_{1,xj}/h_1,\underline{v})d\underline{v}\Big\{1+O\Big(\Big\{\frac{\log n}{nh^{d-1}}\Big\}^{1/2}\Big)\Big\}\end{aligned}$$

uniformly in $1\leq i\leq n$. Therefore, the leading term in the asymptotic variance of $\phi_{n1}(x_1)-\phi_1(x_1)$ is the variance of the following term

$$(nh_1)^{-1}e_1\sum_{j=1}^{n}\varphi(\varepsilon_j)S_{np}^{-1}(x_1,\underline{X}_{2j})f_2(\underline{X}_{2j})\int_{[0,1]^{\otimes d-1}}\{K\mu\}(X_{1,xj}/h_1,\underline{v})d\underline{v},$$

which is asymptotically

$$(nh_1)^{-1}\Big\{\int_{[0,1]^{\otimes d-1}}\{fg^2\}^{-1}(x_1,\underline{X}_2)f_2^2(\underline{X}_2)\sigma^2(x_1,\underline{X}_2)d\underline{X}_2\Big\}e_1S_p^{-1}K_2K_2^{\top}S_p^{-1}e_1^{\top}. \tag{60}$$

If $\rho(y;\theta)=(2q-1)(y-\theta)+|y-\theta|$ and $\varphi(\theta)=2qI\{\theta>0\}+(2q-2)I\{\theta<0\}$, we have $g(\underline{x})=2f_{\varepsilon}(0|\underline{x})$ and

$$\sigma^2(\underline{x})=E[\varphi^2(\varepsilon)|\underline{X}=\underline{x}]=4q^2(1-F_{\varepsilon}(0))+4(1-q)^2F_{\varepsilon}(0)=4q(1-q),$$

which when substituted into (60), yields the asymptotic variance for the quantile regression estimator,

$$\tilde{\sigma}^2(x_1)=q(1-q)\Big\{\int_{[0,1]^{\otimes d-1}}f^{-1}(x_1,\underline{X}_2)f_{\varepsilon}^{-2}(0|x_1,\underline{X}_2)f_2^2(\underline{X}_2)d\underline{X}_2\Big\}e_1S_p^{-1}K_2K_2^{\top}S_p^{-1}e_1^{\top}. \qquad \square$$

The next Lemma is due to Davydov (Hall and Heyde (1980), Corollary A.2).

Lemma 5.3 *Suppose that X and Y are random variables which are $\mathcal{G}-$ and $\mathcal{H}-$ measurable, respectively, and that $E|X|^p<\infty$, $E|Y|^q<\infty$, where $p,\ q>1,\ p^{-1}+q^{-1}<1$. Then*

$$|EXY-EXEY|\leq 8\|X\|_p\|Y\|_q\Big\{\sup_{A\in\mathcal{G},B\in\mathcal{H}}|P(AB)-P(A)P(B)|\Big\}^{1-p^{-1}-q^{-1}}.$$

The next lemma is a generalization of some results in the proof of Theorem 2 in Masry (1996).

Lemma 5.4 *Suppose $\{Z_i\}_{i=1}^{\infty}$ is a zero-mean strictly stationary processes with strongly mixing coefficient $\gamma[k]$, and that $|Z_i| \le B_1$, $\sum_{i=1}^n EZ_i^2 + \sum_{i<j} |\mathrm{Cov}(Z_i, Z_j)| \le B_2$. Then for any $\eta > 0$ and integer series $r_n \to \infty$, if $nB_1/\eta \to \infty$ and $q_n \equiv [n/r_n] \to \infty$, we have*

$$P(|\sum_{i=1}^n Z_i| \ge \eta) \le 4\exp\{-\frac{\lambda_n \eta}{4} + \lambda_n^2 B_2\} + C\Psi(n),$$

where $\Psi(n) = q_n\{nB_1/\eta\}^{1/2}\gamma[r_n]$, $\lambda_n = 1/\{2r_n B_1\}$.

Proof. We partition the set $\{1, \cdots, n\}$ into $2q \equiv 2q_n$ consecutive blocks of size $r \equiv r_n$ with $n = 2qr + v$ and $0 \le v < r$. Write

$$V_n(j) = \sum_{i=(j-1)r+1}^{jr} Z_i, \ j = 1, \cdots, 2q$$

and

$$W_n' = \sum_{j=1}^q V_n(2j-1), \ W_n'' = \sum_{j=1}^q V_n(2j), \ W_n''' = \sum_{i=2qr+1}^n Z_i.$$

Then $W_n \equiv \sum_{i=1}^n Z_i = W_n' + W_n'' + W_n'''$. The contribution of W_n''' is negligible as it consists of at most r terms compared of qr terms in W_n' or W_n''. Then by the stationarity of the processes, for any $\eta > 0$,

$$P(W_n > \eta) \le P(W_n' > \eta/2) + P(W_n'' > \eta/2) = 2P(W_n' > \eta/2). \tag{61}$$

To bound $P(W_n' > \eta/2)$, using recursively Bradley's Lemma, we can approximate the random variables $V_n(1), V_n(3), \cdots, V_n(2q-1)$ by independent random variables $V_n^*(1), V_n^*(3), \cdots, V_n^*(2q-1)$, which satisfy that for $1 \le j \le q$, $V_n^*(2j-1)$ has the same distribution as $V_n(2j-1)$ and

$$P\Big(|V_n^*(2j-1) - V_n(2j-1)| > u\Big) \le 18(\|V_n(2j-1)\|_\infty/u)^{1/2} \sup|P(AB) - P(A)P(B)|, \tag{62}$$

where u is any positive value such that $0 < u \le \|V_n(2j-1)\|_\infty < \infty$ and the supremum is taken over all sets of A and B in the $\sigma-$algebras of events generated by $\{V_n(1), V_n(3), \cdots, V_n(2j-3)\}$ and $V_n(2j-1)$ respectively. By the definition of $V_n(j)$, we can see that $\sup|P(AB) - P(A)P(B)| = \gamma[r_n]$. Write

$$\begin{aligned} P(W_n' > \frac{\eta}{2}) \ &\le P\Big(\Big|\sum_{j=1}^q V_n^*(2j-1)\Big| > \frac{\eta}{4}\Big) + P\Big(\Big|\sum_{j=1}^q V_n(2j-1) - V_n^*(2j-1)\Big| > \frac{\eta}{4}\Big) \\ &\equiv I_1 + I_2. \end{aligned} \tag{63}$$

We bound I_1 as follows. Let $\lambda = 1/\{2B_1 r\}$. Since $|Z_i| \leq B_1$, $\lambda|V_n(j)| \leq 1/2$, then using the fact that $e^x \leq 1 + x + x^2/2$ holds for $|x| \leq 1/2$, we have

$$E\Big\{e^{\pm\lambda V_n^*(2j-1)}\Big\} \leq 1 + \lambda^2 E\{V_n(j)\}^2 \leq e^{\lambda^2 E\{V_n^*(2j-1)\}^2}. \tag{64}$$

By Markov inequality, (64) and the independence of the $\{V_n^*(2j-1)\}_{j=1}^q$, we have

$$\begin{aligned}
I_1 &\leq e^{-\lambda\eta/4}\Big[E\exp\Big(\lambda\sum_{j=1}^q V_n^*(2j-1)\Big) + E\exp\Big(-\lambda\sum_{j=1}^q V_n^*(2j-1)\Big)\Big] \\
&\leq 2\exp\Big(-\lambda\eta/4 + \lambda^2\sum_{j=1}^q E\{V_n^*(2j-1)\}^2\Big) \\
&\leq 2\exp\Big\{-\lambda\eta/4 + C_2\lambda^2 B_2\Big\}. \qquad (65)
\end{aligned}$$

We now bound the term I_2 in (63). Notice that

$$I_2 \leq \sum_{j=1}^q P\Big(\Big|V_n(2j-1) - V_n^*(2j-1)\Big| > \frac{\eta}{4q}\Big).$$

If $\|V_n(2j-1)\|_\infty \geq \eta/(4q)$, substitute $\eta/(4q)$ for u in (62),

$$I_2 \leq 18q\{\|V_n(2j-1)\|/\eta/(4q)\}^{1/2}\gamma[r_n] \leq Cq^{3/2}/\eta^{1/2}\gamma[r_n](r_n B_1)^{1/2}, \tag{66}$$

If $\|V_n(2j-1)\|_\infty < \eta/(4q)$, let $u \equiv \|V_n(2j-1)\|_\infty$ in (62) and we have

$$I_2 \leq Cq\gamma[r_n],$$

which is of smaller order than (66), if $nB_1/\eta \to \infty$. Thus by (61), (63), (65) and (66),

$$P(W_n > \eta) \leq 4\exp\{-\lambda_n\eta/4 + C_2 B_2\lambda_n^2\} + C\Psi_n,$$

where the constant C is independent of n. $\square$

Lemma 5.5 *For any $\underline{x} \in R^d$, let $\psi_{\underline{x}}(\underline{X}_i, Y_i) = I(|\underline{X}_{ix}| \leq h)\psi_x(\underline{X}_{ix}, Y_i)$, a measurable function of $(\underline{X}_i, Y_i)$ with $|\psi_{\underline{x}}(\underline{X}_i, Y_i)| \leq B$ and $V = E\psi_{\underline{x}}^2(\underline{X}_i, Y_i)$. Suppose the mixing coefficient $\gamma[k]$ satisfies (20). Then*

$$\mathrm{Cov}(\sum_{i=1}^n |\psi_{\underline{x}}(\underline{X}_i, Y_i)|) = nV\Big[1 + o\Big\{\Big(B^2 h^{p+d+1}/V\Big)^{1-2/\nu_2}\Big\}\Big].$$

Proof. Denote $\psi_{\underline{x}}(\underline{X}_i, Y_i)$ by ψ_{ix}. First note that

$$V = E\psi_{ix}^2 = h^d \int_{|\underline{u}|\le 1} E(\psi_{ix}^2|\underline{X}_i = \underline{x} + h\underline{u})f(\underline{x} + h\underline{u})d\underline{u},$$

$$\begin{aligned}\sum_{i<j} |\mathrm{Cov}(\psi_{ix}, \psi_{jx})| &= \sum_{l=1}^{n-d}(n-l-d+1)|\mathrm{Cov}(\psi_{0x}, \psi_{lx})| \le n\sum_{l=1}^{n-d}|\mathrm{Cov}(\psi_{0x}, \psi_{lx})| \\ &= n\sum_{l=1}^{d-1} + n\sum_{l=d}^{\pi_n} + n\sum_{l=\pi_n+1}^{n-d} \equiv nJ_{21} + nJ_{22} + nJ_{23},\end{aligned}$$

where $\pi_n = h^{(p+d+1)(2/\nu_2-1)/a}$. For J_{21}, there might be an overlap between the components of $\underline{X}_0$ and $\underline{X}_l$, for example, when $\underline{X}_i = (X_{i-d}, \cdots, X_{i-1})$, where $\{X_i\}$ is a univariate time series. Without loss of generality, let $\underline{u}', \underline{u}''$ and $\underline{u}'''$ of dimensions $l, d-l$ and l respectively, be the $d+l$ distinct random variables in $(\underline{X}_{0x}/h, \underline{X}_{lx}/h)$. Write $\underline{u}_1 = (\underline{u}'^{\top}, \underline{u}''^{\top})^{\top}$ and $\underline{u}_2 = (\underline{u}''^{\top}, \underline{u}'''^{\top})^{\top}$. Then by Cauchy inequality, we have

$$\left|E\left(\psi_{0x}, \psi_{lx}\Big|{}^{\underline{X}_0 = \underline{x} + h\underline{u}_1}_{\underline{X}_l = \underline{x} + h\underline{u}_2}\right)\right| \le \left\{E(\psi_{0x}^2|\underline{X}_0 = \underline{x} + h\underline{u}_1)E(\psi_{jx}^2|\underline{X}_j = \underline{x} + h\underline{u}_2)\right\}^{1/2} = V/h^d \quad (67)$$

and through a transformation of variables, we have

$$|\mathrm{Cov}(\psi_{0x}, \psi_{lx})| \le h^l V \int_{\substack{|\underline{u}_1| \le 1 \\ |\underline{u}_2| \le 1}} |f(\underline{x} + h\underline{u}_1, \underline{x} + h\underline{u}_2; l) - f(\underline{x} + h\underline{u}_1)f(\underline{x} + h\underline{u}_2; l+d-1)|d\underline{u}'d\underline{u}''d\underline{u}''',$$

where by (A4) and (A5), the integral is bounded. Therefore,

$$nJ_{21} \le CnV\sum_{l=1}^{d-1} h^l = o(nV).$$

For J_{22}, there is no overlap between the components of $\underline{X}_0$ and $\underline{X}_l$. Let $\underline{X}_{0x} = h\underline{u}$ and $\underline{X}_{lx} = h\underline{v}$ and we have

$$\begin{aligned}|\mathrm{Cov}(\psi_{0x}, \psi_{lx})| &\le h^{2d} \int_{\substack{|\underline{u}| \le 1 \\ |\underline{v}| \le 1}} E\left(\psi_{0x}, \psi_{lx}\Big|{}^{\underline{X}_0 = \underline{x} + h\underline{u}}_{\underline{X}_l = \underline{x} + h\underline{v}}\right)d\underline{u}d\underline{v} \\ &\qquad \times[f(\underline{x} + h\underline{u}, \underline{x} + h\underline{v}; l+d-1) - f(\underline{x} + h\underline{u})f(\underline{x} + h\underline{v})] \\ &= Ch^d V,\end{aligned}$$

where the last equality follows from (A4), (A5) and (67). Therefore, as $\pi_n h^d \to 0$,

$$nJ_{22} = O\{n\pi_n h^d V\} = o(nV).$$

For J_{23}, using Davydov's lemma (Lemma 5.3) we have

$$|\mathrm{Cov}(\psi_{0x}, \psi_{lx})| \leq 8\{\gamma[l-d+1]\}^{1-2/\nu_2}\{E|\psi_{ix}|^{\nu_2}\}^{2/\nu_2}, \quad \text{as } \nu_2 > 2. \tag{68}$$

As $|\psi_{ix}| \leq B$, $E|\Phi_{ni}|^{\nu_2} \leq B^{\nu_2-2}V$,

$$J_{23} \leq CB^{(\nu-2)2/\nu_2}V^{2/\nu_2}/\pi_n^a \sum_{l=\pi_n+1}^{\infty} l^a\{\gamma[l-d+1]\}^{1-2/\nu_2},$$

where the summation term is $o(1)$ as $\pi_n \to \infty$. Thus $J_{23} = o\Big\{V\Big(B^2h^{p+d+1}/V\Big)^{1-2/\nu_2}\Big\}$, which completes the proof. □

Lemma 5.6 *Suppose (A2)- (A6) hold. Then for $U_{ni}^l, l = 1, \cdots, m$ defined in (50) and $Z_{ni}, l = 1, \cdots, \mathrm{L}_n$ defined in (31), we have*

$$\sum_{i=1}^{n} E(U_{ni}^l)^2 + \sum_{i<j} |\mathrm{Cov}(U_{ni}^l, U_{nj}^l)| \leq Cnh^d M_n^{(1)}\{M_n^{(2)}/M_n^{(1)}\}^{1-2/\nu_2}, \tag{69}$$

$$\sum_{i=1}^{n} EZ_{ni}^2 + \sum_{i<j} |\mathrm{Cov}(Z_{ni}, Z_{nj})| = nh^d(M_n^{(1)})^2 M_n^{(2)}\{M^l \log n\}^{-2/\nu_2}, \tag{70}$$

uniformly in $\underline{x}_k$, $1 \leq k \leq \mathrm{T}_n$.

Proof. We only prove (70), which is more involved than (69). To simplify the notations, denote $\alpha_{j_l}, \beta_{k_l}, \alpha_{j_l}$ and β_{j_l} by $\alpha_1, \beta_1, \alpha_2$ and β_2, respectively. Clearly,

$$\int_{\underline{u}^\top H_n\beta_2}^{\underline{u}^\top H_n(\alpha_2+\beta_2)} \{\varphi_{ni}(\underline{x}_k; t) - \varphi_{ni}(\underline{x}_k; 0)\}dt = \int_{\underline{u}^\top H_n\beta_1}^{\underline{u}^\top H_n(\alpha_2+\beta_1)} \{\varphi_{ni}(\underline{x}_k; t+\underline{u}^\top H_n(\beta_2-\beta_1)) - \varphi_{ni}(\underline{x}_k; 0)\}dt,$$

and

$$\begin{aligned} Z_{ni} &= \int_{\underline{u}^\top H_n\beta_1}^{\underline{u}^\top H_n(\alpha_1+\beta_1)} \{\varphi_{ni}(\underline{x}_k; t) - \varphi_{ni}(\underline{x}_k; 0)\}dt - \int_{\underline{u}^\top H_n\beta_2}^{u^\top H_n(\alpha_2+\beta_2)} \{\varphi_{ni}(\underline{x}_k; t) - \varphi_{ni}(\underline{x}_k; 0)\}dt \\ &= \int_{\underline{u}^\top H_n\beta_1}^{\underline{u}^\top H_n(\alpha_1+\beta_1)} \{\varphi_{ni}(\underline{x}_k; t) - \varphi_{ni}(\underline{x}_k; t + \underline{u}^\top H_n(\beta_2-\beta_1))\}dt \\ &\quad - \int_{\underline{u}^\top H_n(\alpha_1+\beta_1)}^{\underline{u}^\top H_n(\alpha_2+\beta_1)} \{\varphi_{ni}(\underline{x}_k; t + \underline{u}^\top H_n(\beta_2-\beta_1)) - \varphi_{ni}(\underline{x}_k; 0)\}dt \equiv \Delta_1 + \Delta_2. \end{aligned}$$

Therefore, $E\{Z_{ni}\}^2 = h^d \int K^2(\underline{u}) f(\underline{x}_k + h\underline{u}) E\{(\Delta_1 + \Delta_2)^2 | X_i = \underline{x}_k + h\underline{u}\} d\underline{u}$. The conclusion is thus obvious observing that by Cauchy inequality and (22),

$$
\begin{aligned}
E(\Delta_1^2 | X_i = \underline{x}_k + h\underline{u}) &\leq |\underline{u}^\top H_n \alpha_1 \underline{u}^\top H_n (\beta_2 - \beta_1) \underline{u}^\top H_n \alpha_1| \leq 2(M_n^{(1)})^2 M_n^{(2)} / (M^l \log n), \\
E(\Delta_2^2 | X_i = \underline{x}_k + h\underline{u}) &\leq \{\underline{u}^\top H_n (\alpha_2 - \alpha_1)\}^2 (|\underline{u}^\top H_n \alpha_2| + |\underline{u}^\top H_n \alpha_1| + 2|\underline{u}^\top H_n \beta_2|) \\
&\leq 4(M_n^{(1)})^2 M_n^{(2)} / (M^l \log n)^2,
\end{aligned}
$$

where we used the facts that $|\alpha_1 - \alpha_2| \leq 2M_n^{(1)}/(M^l \log n)$ and $|\beta_1 - \beta_2| \leq 2M_n^{(2)}/(M^l \log n)$. Therefore, $E\{Z_{ni}\}^2 = Ch^d (M_n^{(1)})^2 M_n^{(2)}/(M^l \log n)$. As $|Z_{ni}| \leq CM_n^{(1)}$ and $h^{p+1}/M_n^{(2)} < \infty$, the rest of the proof can be completed following the proof of Lemma 5.5. □

Lemma 5.7 *Suppose (A2)- (A6) hold.*

$$
\sum_{i=1}^n E\Phi_{ni}^2 + \sum_{i<j} |\mathrm{Cov}(\Phi_{ni}, \Phi_{nj})| \leq Cnh^d (M_n^{(1)})^2 M_n^{(2)}, \tag{71}
$$

uniformly in $\underline{x} \in \mathcal{D}, \alpha \in B_n^{(1)}$ and $\beta \in B_n^{(2)}$.

Proof. By Cauchy inequality and (22), we have

$$
\begin{aligned}
& E\Phi_{ni}^2 \\
=& h^d \int K^2(\underline{u}) E\Big[\Big\{ \int_{\mu(\underline{u})^\top H_n \beta}^{\mu(\underline{u})^\top H_n(\alpha+\beta)} \Big(\varphi_{ni}(\underline{x}; t) - \varphi_{ni}(\underline{x}; 0)\Big) dt \Big\}^2 | \underline{X}_i = \underline{x} + h\underline{u} \Big] f(\underline{x} + h\underline{u}) d\underline{u} \\
\leq& h^d \int f(\underline{x} + h\underline{u}) K^2(\underline{u}) \mu(\underline{u})^\top H_n \alpha \int_{\underline{u}^\top H_n \beta}^{\mu(\underline{u})^\top H_n(\alpha+\beta)} E\Big[\Big(\varphi_{ni}(\underline{x}; t) - \varphi_{ni}(\underline{x}; 0)\Big)^2 | \underline{X}_i = \underline{x} + h\underline{u}\Big] dt d\underline{u} \\
\leq& h^d \int K^2(\underline{u}) \mu(\underline{u})^\top H_n \alpha \int_{\mu(\underline{u})^\top H_n \beta}^{\mu(\underline{u})^\top H_n(\alpha+\beta)} C|t| dt f(\underline{x} + h\underline{u}) d\underline{u} = O\Big\{ h^d (M_n^{(1)})^2 M_n^{(2)} \Big\},
\end{aligned} \tag{72}
$$

uniformly in $\underline{x} \in \mathcal{D}$, $\alpha \in B_n^{(1)}$ and $\beta \in B_n^{(2)}$. (71) thus follows from (72) and Lemma 5.5. □

Lemma 5.8 *Let* $(A3) - (A6)$ *hold. Then*

$$
\sup_{\underline{x} \in \mathcal{D}} |S_{np}(\underline{x}) - g(\underline{x}) f(\underline{x}) S_p| = O(h + (nh^d/\log n)^{-1/2}) \ \textit{almost surely.}
$$

Proof. The result is almost the same as Theorem 2 in Masry (1996). Especailly if (21) holds, then the requirement (3.8a) there on the mixing coefficient $\gamma[k]$ is met. □

Lemma 5.9 *Denote* $d_{n1} = (nh^d)^{1-\lambda_1-2\lambda_2}(\log n)^{\lambda_1+2\lambda_2}$ *and let* λ_1 *and* $B_n^{(i)}$, $i = 1, 2$, *be as in Lemma 5.1. Suppose that* (A1) – (A5) *and (19) hold. Then there is a constant* $C > 0$ *such that for each* $M > 0$ *and all large* n,

$$\sup_{\underline{x}\in\mathcal{D}} \sup_{\substack{\alpha \in B_n^{(1)},\\ \beta \in B_n^{(2)}}} |\sum_{i=1}^{n} E\Phi_{ni}(\underline{x};\alpha,\beta) - \frac{nh^d}{2}(H_n\alpha)^\top S_{np}(\underline{x})H_n(\alpha+2\beta)| \leq CM^{3/2}d_{n1}.$$

Proof. Recall that $G(t,\underline{u}) = E(\varphi(Y;t)|\underline{X} = \underline{u})$,

$$\begin{aligned} E\Phi_{ni}(\underline{x};\alpha,\beta) \quad = \quad & h^d \int K(\underline{u})f(\underline{x}+h\underline{u})d\underline{u} \times \int_{\mu(\underline{u})^\top H_n\beta}^{\mu(\underline{u})^\top H_n(\alpha+\beta)} \\ & \Big\{G(t+\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u}) - G(\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u})\Big\}dt. \end{aligned} \tag{73}$$

By $(A3)$ and $(A5)$, we have

$$\begin{aligned} & G(t+\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u}) - G(\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u}) \\ & \qquad = tG_1(\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u}) + \frac{t^2}{2}G_2(\xi_n(t,\underline{u};\underline{x}), \underline{x}+h\underline{u}), \\ & G_1(\mu(\underline{u})^\top H_n\beta_p(\underline{x}), \underline{x}+h\underline{u}) = g(\underline{x}+h\underline{u}) + O(h^{p+1}), \end{aligned}$$

where $\xi_n(t,\underline{u};\underline{x})$ falls between $\mu(\underline{u})^\top H_n\beta_p(\underline{x})$ and $t+\mu(\underline{u})^\top H_n\beta_p(\underline{x})$, and the term $O(h^{p+1})$ is uniform in $\underline{x} \in \mathcal{D}$. Therefore, the inner integral in (73) is given by

$$\frac{1}{2}g(\underline{x}+h\underline{u})(H_n\alpha)^\top \mu(\underline{u})\mu(\underline{u})^\top H_n(\alpha+2\beta) + O\Big\{M^{3/2}\Big(\frac{\log n}{nh^d}\Big)^{\lambda_1+2\lambda_2}\Big\}$$

uniformly in $\underline{x} \in \mathcal{D}$, where we have used the fact that $nh^{d+(p+1)/\lambda_2}/\log n < \infty$. By the definition of $S_{np}(\underline{x})$, the proof is thus completed. □

Lemma 5.10 *Under conditions in Theorem 3.2, we have*

$$\sup_{\underline{x}\in\mathcal{D}} \Big|\frac{1}{nh^d}W_pS_{np}^{-1}(\underline{x})H_n^{-1}\sum_{i=1}^{n}K_h(\underline{X}_i-\underline{x})\varphi(\varepsilon_i)\mu(\underline{X}_i-\underline{x})\Big| = O\Big\{\Big(\frac{\log n}{nh^d}\Big)^{1/2}\Big\} \text{ almost surely.}$$

Proof. Note that, under conditions Theorem 3.2, the assumptions imposed by Masry (1996) in Theorem 5 are validated. Specifically, (4.5) there follows from (19) and (4.7b) there can be derived from (21). Therefore, following the proof lines there, we can show that

$$\sup_{\underline{x}\in\mathcal{D}} \Big|\frac{1}{nh^d}H_n^{-1}\sum_{i=1}^{n}K_h(\underline{X}_i-\underline{x})\varphi(\varepsilon_i)\mu(\underline{X}_i-\underline{x})\Big| = O\Big\{\Big(\frac{\log n}{nh^d}\Big)^{1/2}\Big\},$$

which together with Lemma 5.8 yields the desired results. □

REFERENCES

Andrews, D.W.K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62**, 43-72.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577-80.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. NewYork: Springer-Verlag.

Chen, X., Linton, O. B. and I. Van Keilegom (2003). Estimation of Semiparametric Models when the Criterion is not Smooth. *Econometrica* **71**, 1591-608.

Fan, J., Heckman, N.E. and Wand, M.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90**, 141-50.

Fan, J. and Gijbels, I. (1996). *Local polynomial regression.* London: Chapman and Hall.

Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.* **95**, 246 - 72.

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications*. NewYork: Academic Press.

Hong, S. (2003). Bahadur representation and its application for Local Polynomial Estimates in Nonparametric M-Regression. *J. Nonparametric Statist.* **15**, 237-51.

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100**, 1238-49.

Huber, P. J. (1973) Robust regression. *Ann. Statist.* **1**, 799-821.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38**, 1323-42.

Linton, O. B. (2001). Estimating additive nonparametric models by partial L_q Norm: The Curse of Fractionality. *Econom. Theory* **17**, 1037-50.

Linton, O. B., Hardle, W and Sperlich, S (1999). A Simulation comparison between the Backfitting and Integration methods of estimating Separable Nonparametric Models. *TEST* **8**, 419-58.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-100.

Linton, O. B., Sperlich, S. and I. Van Keilegom (2007). Estimation of a Semiparametric Transformation Model by Minimum Distance. *Ann. Statist.* To appear

Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika* **83**, 529-40.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Ser. Anal.* **17**, 571-99.

Peng, L. and Yao, Q. (2003). Least absolute deviation estimation for ARCH and GARCH models. *Biometrika* **90**, 967-75.

Rosenblatt, M. A central limit theorem and strong mixing conditions. *Prof Nat. Acad. Sci.* **4**, 43-7.

Sperlich, S., O. Linton, and W. Härdle (1998). A Simulation comparison between the Backfitting and Integration methods of estimating separable nonparametric models. *Test* **8**, 419-58.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-53.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 592-606.

Wu, W. B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33**, 1934-63.